From short-term goals to organizational objectives:

Managing security control using goal setting

Thesis

presented to the Faculty of Arts

of the University of Zurich

for the degree of Doctor of Philosophy

by

Olive Emil Wetter

of Remetschwil (AG), Switzerland

Accepted in the spring semester 2012 on the recommendation of

Prof. Dr. Klaus Jonas and Prof. Dr. Alexandra Freund

Zurich, Switzerland, 2012

Abstract

This PhD thesis features the first three studies on the application of goal setting at an airport's security control. Goal setting is one of the most studied areas in work and organizational psychology. The first study analyzed if goal setting by supervisors can easily and effectively be used to guide work performance of Security Officers at a security control. The results of this field study demonstrate that the use of short-term goals is practicable even without prior investments such as training of supervisors. Furthermore, an effect of goal setting on objective team-level measures of productivity could be demonstrated. The second study is a laboratory study that contrasted a dual goal for speed and accuracy with a single goal for speed on the same task. As performance tasks, the Sternberg paradigm and the d2 test were used. Speed measures and errors demonstrated that both dual and single goals increase performance by enhancing memory scanning. The single goal triggered a speed-accuracy trade-off, favoring speed over accuracy, whereas this was not the case with the dual goal. In difficult trials, dual goals slowed down scanning processes again so that errors could be prevented. The third study analyzed goals on the highest hierarchical level, namely organizational level. Using the Productivity Measurement and Enhancement System, the relevant organizational objectives, goal domains, as well as the key performance indicators of a security control could be identified. The operationalizations of the key performance indicators allow for insights into measurement practices at a security control. The contingency functions of the system allow managers to judge in which domain investments are most promising.

Abstract (Deutsch)

Diese Doktorarbeit beinhaltet die ersten drei Studien, die die Anwendung von Zielsetzungen in einer Sicherheitskontrolle an einem Flughafen thematisieren. Zielsetzungen sind eines der am meisten untersuchten Gebiete innerhalb der Arbeits- und Organisationspsychologie. Die erste Studie untersuchte, ob Zielsetzungen durch Vorgesetzte einfach und effektiv verwendet werden können, um die Arbeitsleistung von Sicherheitsbeauftragten einer Sicherheitskontrolle zu lenken. Die Resultate dieser Feldstudie zeigen, dass Ziele, die eine kurze Zeitspanne betreffen, sogar ohne vorgängige Investitionen wie zum Beispiel Schulung der Vorgesetzten praktikabel sind. Darüber hinaus konnte ein Effekt der Zielsetzungen auf objektive Produktivitätsmasse auf Teamebene gezeigt werden. Die zweite Studie ist eine Laborstudie, welche ein Doppelziel für Geschwindigkeit und Genauigkeit mit einem Einfachziel für Geschwindigkeit bezüglich derselben Aufgabe verglich. Als Leistungsaufgaben wurden das Sternberg-Paradigma und der d2 Test verwendet. Geschwindigkeitsmasse und Fehler zeigten auf, dass sowohl Doppel- als auch Einfachziele die Leistung durch Verbesserung des Scannings im Gedächtnis erhöhen. Das Einfachziel löste einen Geschwindigkeits-Genauigkeits-Austausch aus, wobei Geschwindigkeit gegenüber Genauigkeit begünstigt wurde. Mit dem Doppelziel war dies nicht der Fall. In schwierigen Experimentaldurchgängen verlangsamten Doppelziele die Scanningprozesse wiederum, was die Vermeidung von Fehlern ermöglichte. Die dritte Studie analysierte Ziele auf der höchsten Hierarchiestufe, nämlich auf Organisationsebene. Durch die Anwendung des Productivity Measurement and Enhancement System konnten die relevanten Organisationsziele, Zielbereiche, sowie die Leistungsindikatoren einer Sicherheitskontrolle identifiziert werden. Die Operationalisierungen der Leistungsindikatoren ermöglichen Einblicke in die Leistungsmessung einer Sicherheitskontrolle. Die Kontingenzfunktionen des Systems ermöglichen es Managern abzuschätzen, in welchem Bereich Investitionen am lohnenswertesten erscheinen.

Table of contents

1 General introduction

1.1 The history of research and development in airport security

Airport security is a relatively young subject as such, and its history within psychology is even shorter. The terror attacks against the United States of America on September 11, 2001, can be identified as the major driving force which boosted both airport security industry and research. 9/11 as well as later incidents, which received broad coverage in mass media, drastically increased both the public and governmental interest as well as the perceived need for research in this field. Politicians, government authorities, and the general public alike called for more stringent security measures that had to be devised and subsequently evaluated. In many cases, those had a considerable impact on society.

For example, the stringent restrictions concerning liquids in cabin baggage[1] (European Commission, 2006) for airline passengers after the 2006 transatlantic aircraft plot, where liquid components of explosives were planned to be smuggled on board an aircraft (Srinivasan & Prabhakaran, 2009), considerably restricted passengers' personal freedom. However, at that time, it was the only possible way of mitigating the threat posed by liquid explosives. In everyday practice though, this threat mitigation came at a high price. According to the theory of reactance (Brehm, 1966; Brehm & Brehm, 1981; Miron & Brehm, 2006), the greater the subjectively perceived importance of a restricted personal freedom is, the more reactance the individual will show. As drinking is the most important vital requirement for human beings, it is not astonishing that reactivity was rather strong. The ensuing problems at airport security control ranged from tedious discussions with passengers and innumerable complaints to physical attacks on police staff. Apart from the annoyance caused for passengers, this new restriction also meant significant additional investments in terms of time and

[1] “Cabin baggage” (also known as “carry-on baggage/luggage” or “hand baggage/luggage”) is the type of baggage that a passenger is allowed to carry along in the cabin of the aircraft.

money, which again stressed the importance of research and development in this field in order to alleviate these negative consequences as much as possible.

There are basically two main areas where investments into research and development can be made; namely technology and human factors. In the first years after 9/11, measures taken in order to render flying more secure usually consisted of large investments into technology. Training and subsequent use of the human capabilities of Security Officers (SOs) were of secondary importance. Still today, some authors stress the importance of new technology as an area of potential improvement (e.g., Seidenstat, 2009). As one of the first researchers worldwide, Adrian Schwaninger realized this disequilibrium of technology and the human factor. In order to make the human potential accessible, he (Schwaninger, 2005) combined signal detection and object recognition theories, which led to the development of a pre-employment assessment test for x-ray screeners (x-ray object recognition test ORT, Hardmeier, Hofer, & Schwaninger, 2005), computer-based training for x-ray image interpretation (X-Ray Tutor, Schwaninger, 2003, 2004), as well as the x-ray competency assessment test (X-Ray CAT, Koller & Schwaninger, 2006).

In the following years, many studies validated the existing approaches, as well as broadened and extended the available knowledge in these domains. Interestingly, there are only a few studies available from abroad: McCarley, Kramer, Wickens, Vidoni, and Boot (2004) contributed significantly to the development of x-ray screener training. Meneer, Barrett, Phillips, Donnelly, and Cave (2007) investigated in more detail what happens if two targets have to be identified simultaneously in an image. Most studies, however, were carried out by the Schwaninger group. They further investigated which factors influence detection performance in x-ray screening and their relative importance (Bolfing, Halbherr, & Schwaninger, 2008), identified how the relation of search time and non-search time in x-ray baggage inspection changes with training using the two-component inspection model (Koller, Drury, & Schwaninger, 2009; Wales, Anderson, Jones, Schwaninger, & Horne, 2009), and compared three different computer-based training systems (Koller, Hardmeier, Michel, & Schwaninger, 2008).

Once pre-employment assessment, computer-based training, and competency assessment for x-ray screeners had been established, it became evident that performance in training and testing sessions is not necessarily equivalent to everyday performance at the security control lane. Imagine a SO who is perfectly trained and able to detect prohibited items in cabin baggage. If this SO is, for whatever reason, inattentive at work, his performance on the job is likely to be much worse than his performance in training and testing sessions. These thoughts encouraged researchers and the industry alike to think about ways of measuring on the job performance of SOs. The first approach in that domain was realized by Hofer and Schwaninger (2005) who studied data from Threat Image Projection (TIP) in cabin baggage screening.

TIP is a software function of x-ray devices that presents threat images to SOs during work. Threat images are shown with a preset overall frequency and appear at random intervals. The first goal of TIP is to keep SOs' alertness at a high level during work. Without TIP, the usually extremely low prevalence of threat objects encountered by SOs during their daily work can be regarded as a danger to their alertness. Because SOs working without TIP would experience again and again that threats object are extremely rare, their criterion of decision might tend to become more liberal (Wolfe et al., 2007). The second goal of TIP is to allow for measuring on the job performance of x-ray screeners by logging SOs reactions to threat images. Carrying out research with TIP data, however, is somewhat delicate since those data are security sensitive information. Nevertheless, ways have been found to continue research in that domain.

The study by Schwaninger, Hofer, and Wetter (2007) evaluated the adaptive computer-based training system X-Ray Tutor using TIP data. This study allowed for a validation of previous results. Moreover, it examined training and TIP in hold baggage[2] screening (Shanks & Bradley, 2004) more closely, which differs from cabin baggage screening in several points. Hold baggage is larger in size, often packed in a much more complex way than cabin baggage, and the list of prohibited items differs from cabin baggage. Furthermore, x-ray systems for hold baggage are different from x-ray systems for cabin baggage: for instance, they feature automatic explosives detection and generate images from multiple viewpoints. TIP has different properties in hold baggage screening as well: In hold baggage screening, so-called “combined threat/non-threat images” (CTIs/CNTIs) are used, which are previously assembled images, either with or without a threat, that are taken from an image library and inserted in between real passenger bag images. In cabin baggage screening, however, so-called “fictional threat items” (FTIs), that are taken from an image library, are inserted on the spot into x-ray images of real passenger bags.

1.2 The “embedded applied research and development” paradigm

Up to 2007, sophisticated machines and proven ways of training x-ray image interpretation of SOs had been developed. All of this work could be done in the university setting by analyzing automatically generated data. In the same year, the new head of the security control branch of the Kantonspolizei Zürich (Zurich State Police) realized that he had to have a research and development team inside the organization in order to advance the field any further. In 2008, he formed this team and, together with them, created a research setting that he called “embedded applied research and development”. This term signifies that a group of researchers actually works within the organization and is part of it. This paradigm offers the advantage of the researchers being able to have deep insights into the organization and the challenges it faces. The

[2] “Hold baggage” (also known as “checked baggage”) is the type of baggage that a passenger hands over to staff at check-in, which is then transported in the hold of the aircraft.

researchers spend a considerable amount of time learning about the organization and getting to know the workforce of the organization.

The result of the first insights gained with this new paradigm were that it was realized that x-ray screening with training and TIP applications represents just one part of the SOs' work. On being able to scrutinize the daily process at a security control, it became evident that there are in fact many more duties and tasks to be carried out reliably and efficiently. In order to provide a high security standard, not only cabin and hold baggage have to be scrutinized but also the passengers themselves. It became clear that a holistic approach would be needed. A holistic approach tries to account for the whole final product, and security control is seen as composed of several different elements that altogether contribute to this final product. Among these elements are processes like x-ray screening, correct operation and interpretation of the hand-held metal detector and the walk-through metal detector[3], use of trace detection equipment[4], pat-down search[5], manual baggage search, and correct reaction and communication in case of an alarm.

In order to measure performance of a security control in a more holistic way, the method of covert testing was at hand. Wetter, Hardmeier, and Hofer (2008)[6] defined covert testing as “a method of testing the security control’s ability to detect threat objects carried by a tester in a situation of high ecological validity” (p. 358). Covert tests had already been used on a case-by-case basis by regulatory authorities for quality control purposes. In the mentioned publication, the method of covert testing was studied in more detail. This study is, to our knowledge, the first published study

[3] Both hand-held and walk-through metal detectors are used to detect metallic objects on a person which are not allowed in the cabin of the aircraft.

[4] Trace detection equipment is used for detecting traces of explosive and/or inflammable substances in baggage and on passengers.

[5] “Pat-down search” refers to the manual inspection of a passenger in order to check if items are carried on the body or in garments that are not allowed in the cabin of the aircraft.

[6] This work was conducted during the time of my PhD studies. It can be seen as a preliminary study investigating more closely one method of performance measurement. The results of this study and the knowledge gained on covert tests served as a basis for definition and measurement of indicator 3.1 of our ProMES system (see chapter 4.4.B).

worldwide on covert testing. It turned out that frequent covert tests are not only an instrument for assessing performance and vulnerabilities in a holistic way, they can also be used to train and motivate SOs, which is particularly important since the base rate of people carrying forbidden items is, luckily, very low in everyday practice. With this study, the "embedded applied research and development" team was able to make its first contribution to broadening the horizon of research in airport security by making use of their in-depth knowledge of security control operation.

In the following year, two more studies were published. One of them (Hofer, Wetter, Graf, Guo, & Schwaninger, 2009[7]) was conducted in collaboration with the Schwaninger group. This study extended the present knowledge on training by analyzing learning content management systems (a form of e-learning) as a means for (re-)training and communication in the security domain. The second study (Wetter, Laube, & Hofer, 2009[7]) consists of a comparison between two different leadership structures for security control lanes. It compares the traditional structure to a new structure, which had been devised by a practitioner with many years of experience in the field. To our knowledge, this study is the first worldwide that focused on leadership and organizational structures in the security control domain. With this study, the domain of work and organizational psychology was opened in airport security research. Both the first and the second study, given in full text later in this PhD thesis (see chapters 2 and 3), can be seen as a continuation of that work.

Providing security is certainly the main task of a security control. However, there are other aspects that are important as well. Wetter, Lipphardt, and Hofer (2010)[8] stated that "the ideal security control would be effective, cheap, and would not interfere with the operation of the airport" (p. 301). These aspects open an interdisciplinary field, which can only successfully be studied if psychologists, engineers, managers, and

[7] This work was also conducted during the time of my PhD studies.
[8] This work was also conducted during the time of my PhD studies. It is a preliminary study to the first and the third study given in full text in this PhD thesis (see chapters 2 and 4).

security personnel work together and share their individual knowledge. In order to do so, an interdisciplinary working group at Zurich Airport was founded in 2009, which was jointly led by the police's research & development team and by a management representative of Zurich Airport's Terminal Engineering. Participating members from the police were the head of passenger control and one of his supervisors, the head of instruction and support, technical staff, the process owner, a project leader, as well as an employee representation. Participating members from Zurich Airport were a representative of Airport Security and the head of electronic support. On a case by case basis, further staff or management representatives were invited to participate. The group reported to a steering committee of airport and police managers.

To be able to more closely examine the aspects mentioned in the above paragraph, new ways of measurement had to be defined. Some of these concepts (e.g., throughput) had been known for some time already in benchmarking between airports. They are of particular importance for airport terminal capacity planning (Solak, Clarke, & Johnson, 2009). For in-depth study of the security control process, it was necessary to more clearly define and operationalize these measures. The study by Wetter et al. (2010) took that up and proposed definitions and operationalizations of throughput and passenger density. It furthermore specified two different ways of measurement for throughput and also a different method to determine passenger density. Additionally, the study suggested a classification of the influences on the security control process into external and internal factors. By the means of exploratory field work, it estimated the influence of some prominent external and internal factors on objective and subjective dependent variables, covering the domains of efficiency, security, and employee work strain.

1.3 The three "goal setting" studies

In chapters two to four, three studies are given in full text, which represent the continuation of the line of research mentioned above. These three studies have in common that they are all about goal setting, a well-established subject in work and organizational psychology, which harbors an abundance of findings and studies (e.g., Locke & Latham, 2002). The goal of this PhD thesis is to prepare the grounds for applications of goal setting in security related domains, which pose their own specific demands. Furthermore, the effects to be expected by applying goal setting in security related domains shall be analyzed. The scope of the studies becomes broader and more integral from the first to the third, and the picture conveyed more holistic.

The first study (Wetter, Hofer, & Jonas, submitted) has its starting point at the very front: It is a field study that analyzes if goal setting is a practicable way of influencing work performance of SOs by their supervisors. The study is set at an airport's security control and is thus of the highest possible ecological validity. The baseline condition in this study is work without daily goals, i.e., the status quo. As previously used in many goal setting studies, two different goals, namely a “do your best” and a “specific difficult” goal, were used as experimental conditions. With this study, it is demonstrated that daily goal setting is practicable and does indeed make a difference regarding work performance in the security domain. These first positive results are the basis for the two subsequent studies.

During data collection for the first study, some employees voiced their concern about goal setting as we tested it. They argued that speed goals would lead to a decrease in performance in another aspect of their work (e.g., provided security level). In other words, they suggested that trade-offs show up if goals make one focus on one single aspect. These concerns were taken into account, which led us to think about dual and even multiple goals. Would it be possible to tackle several important aspects of one’s work with multiple goals? As a result of these considerations, the second study (Wetter, Wegge, Jonas, & Schmidt, submitted) was designed. It is a laboratory study that investigates the effects of single and dual goals on speed and performance measures. It is guided by the research questions if dual goals are effective, and, if yes, if they have the potential to reduce speed-accuracy trade-offs. The laboratory setting was chosen in order to have as much standardization as possible for investigation of these theoretical issues, as well as not to risk any speed-accuracy trade-off in the real setting at the security control.

Thinking about dual or even multiple goals inevitably leads to the question of what the different goals of the organization actually are. This is the main research question for the third study (Wetter, Fuhrmann, Lipphardt, & Hofer, 2011). It draws upon the knowledge gained in the two preceding studies. By providing the big picture, the subject is wrapped up: The third study identifies goals on the highest hierarchical level in the organization. It demonstrates that, in a complex work domain like an airport, all stakeholders have to work together in order to achieve the best possible result. As can be inferred by the results of the second study, a maximization of performance in every single key performance indicator of the organizational goals is unlikely to be possible due to capacity limits in the human workforce, equipment, as well as in time and resources. Therefore, an optimization has to be sought for. In order to reach the most useful optimization, management level representatives of the two main stakeholders at airport security control set up a Productivity Measurement and Enhancement System (ProMES, Pritchard, 1990) on management level, which specifies clear contingency functions between key performance indicator values and productivity scores. This system provides clear answers if decisions for optimization in one or the other goal domain have to be made. It specifies which key performance indicator in which goal domain has the most potential productivity benefit at that moment if an investment is to be made.

As all three studies are subsequently given in their original text and style, they differ from each other. The first and the second study adhere to the writing standard of the APA (American Psychological Association), as they are submitted to journals that apply the APA standard. The third study, however, adheres to the writing standard of the IEEE (Institute of Electrical and Electronics Engineers), as it was published in a publication of the IEEE. The IEEE is “the world’s largest professional association for the advancement of technology” (IEEE, 2011).

1.4 Theoretical foundation of the three studies

1.4.A Goal setting and the High Performance Cycle[9]

One of the very basic questions in management is how to tell employees what their job consists of and what they are expected to do. On explaining a job profile to subordinates, managers specify tasks to be carried out that should lead to the attainment of certain objectives. In the course of the last century and with the progress in technology, jobs became increasingly complex. It is not astonishing therefore, that optimal ways of communication about duties and expectations at work have been sought for. Goal setting theory specifies in particular, how leaders can effectively communicate their expectations as goals to their subordinates by showing how goals affect performance. The theory further suggests mechanisms by which goals exert their effect, and it specifies mediators and moderators of this process.

The so-called “High Performance Cycle” (Figure 1, see next page) is a framework that explains motivation in the workplace (Latham & Locke, 2007). It is based on the findings of a wide body of studies on goal setting and provides a basis for making interventions. It explains the regulatory cycle that is a complex interplay of various components. The demands include the goal, which shapes subsequent actions. The actions themselves and their result are influenced by moderators and conveyed by mediators. Contingent rewards lead to satisfaction, which again enhances commitment to the organization. To close the cycle, it is postulated that high commitment and willingness to accept future challenges leads to subsequent acceptance of new, challenging, high goals, as well as a high self-efficacy.

[9] As goal setting is one of the most researched subjects in work and organizational psychology with an abundance of studies and publications, only a short and incomplete introduction can be given here.

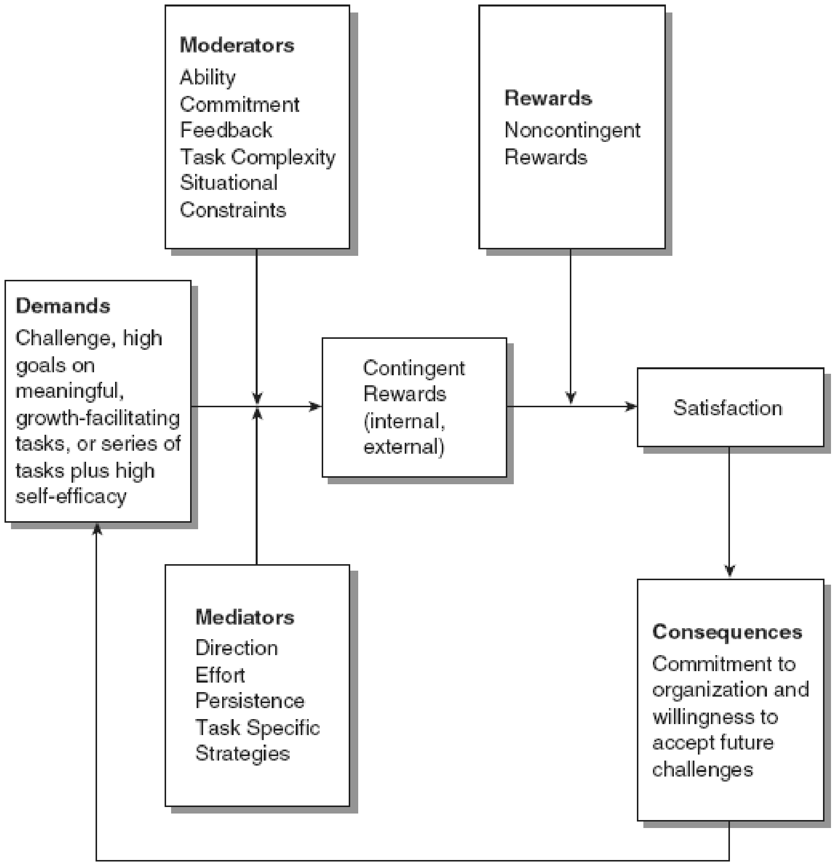

Figure 1. The High Performance Cycle (from Latham & Locke, 2007).

The core of the goal setting theory focuses on the properties of an effective goal, which are goal difficulty and specificity, goal level, content of goals, mediators and moderators, and goal source (Locke & Latham, 2002). As these properties are among the most important tenets of goal setting theory, they shall be defined and discussed in the following paragraphs.

It is proven by many studies that goals that are both specific *and* difficult lead to better performance than unspecific or easy goals, and can thus be regarded as optimal. Easy goals do not trigger much motivation and effort. Unspecific goals (e.g., “do your best”) lead to a great variance in performance. Because they lack an external referent, they can be defined idiosyncratically (Locke & Latham, 2002). They can thus be regarded as being shaped by the person’s motivation and aspiration levels.

"Goal level" refers to the different hierarchical levels on which goals can be set. These range from goals for individuals, which have been researched most, to group and organizational goals, which have moved into the focus of research rather recently. In groups and organizations, the beneficial effects of specific difficult goals can also be observed (Locke & Latham, 1990; Kleingeld, van Mierlo, & Arends, 2011; Widmeyer & Ducharme, 1997). Inside one organization, the picture gets somewhat more complex, though, since individual and group/organizational goals do not necessarily have to match. This can lead to open or hidden goal conflicts. The diagnosis, alignment and management of organizational, group, and individual goals can be seen as one of the major challenges for managers in organizations.

The content of the goals can either focus on performance or on learning. A performance goal consists of a certain level to be attained on a performance measurement indicator (e.g., to produce 100 pieces per hour). A learning goal refers to the number of different strategies, ideas, or problem-solving activities that are developed for a given task. Learning goals can be especially fruitful when people are confronted with a very difficult and complex task (Winters & Latham, 1996). The focus then lies on the process of working on the problem and learning by doing so rather than on the final outcome.

Up to the present, many mediators and moderators of the effects of goal setting have been identified. Probably the most important and obvious moderator is goal commitment. If people are not committed to a goal, there will be no or only small effects of goal setting (Klein, Wesson, Hollenbeck, & Alge, 1999; Seijts & Latham, 2000). Goal commitment is high if the goal is subjectively perceived as important and if self-efficacy, i.e., the belief that the goal can be attained, is high (Locke & Latham, 2002). Task complexity is a second moderator of the effects of goal setting: With tasks of lower complexity, the effects of goal setting are higher than with tasks of high complexity (Wood, Mento, & Locke, 1987). Feedback, which will be discussed in chapter 1.4.C, can be considered as another important moderator of goal commitment. The four main mediators of the effects of goal setting on performance, namely effort, persistence, attention, and task strategies, will be introduced in chapter 3.1.

Last but not least, the goal source can differ as goals can be self-set, participatively set, or assigned. It could be assumed that participatively set goals have a general advantage over assigned goals because participatively set goals might have a higher subjective importance for employees. However, the empirical evidence on this question turned out to be controversial. Locke and Latham (2002) cite several studies by the Latham group which showed that the results reached with participatively set and assigned goals did not differ. On the other hand, the Erez group reached the opposite conclusion. Given these controversial results, both groups decided to design a study together. This approach revealed that it is important that the goal makes sense for the employee (i.e., its purpose or rationale is sensible) in order to be effective, whereas performance will be lower if a goal is assigned and enforced without explanation (Latham, Erez, & Locke, 1988). The former dispute and later collaboration between Latham and Erez underpin the importance of participation as a concept within goal setting, decision-making, and management in general.

1.4.B Participation

Participation is a concept that exists in many different fields of research with different meanings (Wegge, 2004). However, many definitions have in common that participation implies that there is shared decision making, that employees contribute according to their competence and not necessarily according to their position, and that communication channels are open in all directions (Mitchell, 1973). Participation is an important concept because it has been shown to have positive effects on emotions (satisfaction) and motivation of employees (upper part of Figure 2, see next page), as well as beneficial cognitive effects in decision-making (lower part of Figure 2, see next page), which contribute to better performance or productivity (Miller & Monge, 1986).

Figure 2. Positive effects of participation on performance and employee satisfaction (translated from Wegge, 2004).

The positive effects of participation on emotions (satisfaction) and motivation of employees, independent of later performance, can be explained by several mechanisms. Participation helps to satisfy human desires of self-efficacy (Bandura, 1977, 1997) and self-determination (Deci & Ryan, 2002). Moreover, participation helps fulfilling human needs of power, achievement, and affiliation (McClelland, 1987). It can also be assumed that autocratic decisions of managers trigger reactance, which leads to a decrease of positive emotions such as satisfaction or trust (e.g., Sagie, Elizur, & Koslowsky, 1995). By granting participation to subordinates, the negative effects of reactance can be reduced (Ashforth & Sacks, 2000). Last but not least, participation has been shown to also increase goal commitment since the process of taking part in decision-making increases one's feeling of liability (Erez, Earley, & Hulin, 1985; Klein et al., 1999).

Regarding beneficial cognitive effects of participation in decision-making, it could be shown that intelligent sharing of knowledge and information between subordinates and managers leads to better performance (Scully, Kirkpatrick, & Locke, 1995). Managers have to guide the process of information sharing so that they get the relevant information without too much irrelevant or even misleading information. Further beneficial cognitive effects revolve around role clarity. Schuler (1980) postulated that participation leads to less role conflicts at work, which again entails an increase in satisfaction. Smith and Brannick (1990) confirmed Schuler's model empirically and found an additional direct influence of participation on satisfaction apart from the indirect one that had been postulated by Schuler. It is noteworthy, though, that in routine tasks such as carrying out security control, there is presumably not much role ambiguity to be found, so the main cognitive advantages of participation might lie in the sharing of information, knowledge, and experience.

Encouraging and engaging in participation can be beneficial for both managers and subordinates. By granting participation to subordinates, managers have a means of motivating them, increasing their goal commitment, as well as reaching better performance outcomes and better decisions. Subordinates can fulfill some of their needs and desires and can, at least partly, shape their own work. It is important, however, that granted participation also really gives subordinates a chance for participating. “Rhetorical” participation, where subordinates are invited to participate but in the end have no influence at all, will not lead to the postulated beneficial effects. Unfortunately, real, honest participation that is also perceived as such and entails the beneficial effects seems to be rather rare (e.g., Harley, 1999).

1.4.C ProMES and feedback

A systematic approach for setting up a management system is ProMES (Pritchard, 1990). The key to its success is the combination of the beneficial effects of goal setting, participation and empowerment, as well as feedback, which renders it particularly powerful. The effectiveness of ProMES has been proven in many different fields (Pritchard, Harrell, DiazGranados, & Guzman, 2008). Of particular importance for this thesis is certainly the successful application of ProMES in a Swedish traffic police unit (Pritchard, Culbertson, Malm, & Agrell, 2009). One of the keys to the success of ProMES might be its theoretical foundation, which lies in the motivational theory of Naylor, Pritchard, and Ilgen (1980).

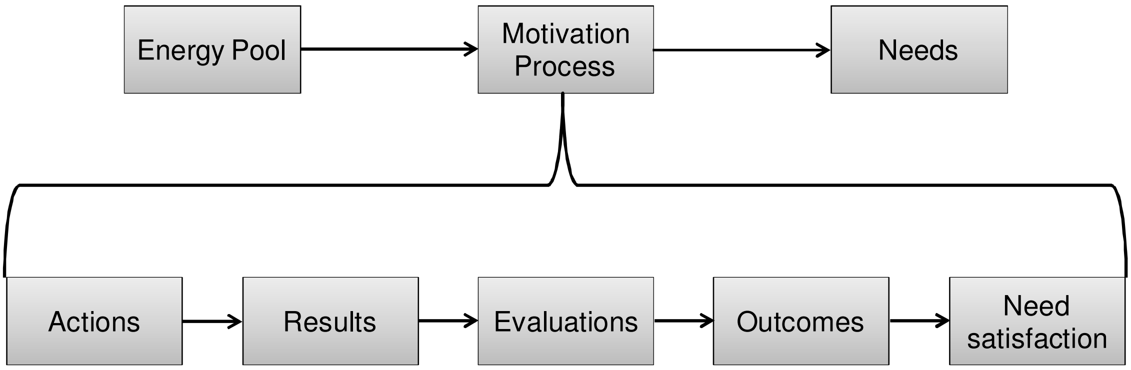

Figure 3. The expanded motivation model (Pritchard & Ashwood, 2008).

The expanded motivation model (Figure 3) states that, on the one hand, there is a pool of physical, mental, and emotional resources ("energy pool") in every person. On the other hand, there are the needs of a person, which are understood in a very broad sense and can include the needs mentioned in chapter 1.4.B. The process of motivation is seen as the expectation of being able to satisfy one's needs by applying resources (energy) to certain actions. Actions generate results that are evaluated subsequently, either by oneself or by another person (e.g., a supervisor). The outcomes can be described as what the person gets back after the results have been evaluated. Depending on what the outcomes consist of, they have more or less potential to satisfy the person's needs.

One of the core tenets of this expectancy-value theory is that the strength of the connections of the components of the motivation process determines the amount of motivation. For example, if actions do not lead to the desired results, the individual's motivation to do the same action will be lower next time and different actions might be chosen to reach the desired results; or if the outcomes do not satisfy the individual's needs, the motivation to strive for the same outcomes will be lower next time. Pritchard and Ashwood (2008) state that "for motivation to be high, all the components of the model must be high" [p. 37].

ProMES can be seen as a system that strives to positively influence the results, evaluations, and outcomes of the motivation process, as well as their connections. A ProMES system specifies and clarifies *which* results will be looked at and evaluated subsequently by defining goal domains and objectives. Contingency functions clearly specify *how* results will be evaluated. By comparing the contingency functions of the different indicators, their relative importance is given. It becomes visible how the evaluation of a single result (productivity score of an indicator) will contribute to the overall outcome for the company (total productivity). In some cases, ProMES was successfully coupled with a reward system that clearly specified the outcome (reward) for the evaluations and let employees benefit from the organization's success.

In order to reach strong motivational effects, it does not suffice that the components are "high" (i.e., in favorable condition) and that the connections between them are strong. In addition, the outcomes have to appear useful to the employees for satisfying their needs. The actual process of satisfying employee needs with the outcomes (e.g., money) is a process that is controlled by the employees themselves. If the motivational chain specified above is clear, understandable, and plausible to employees, their motivation will be high. This will make them choose actions and, subsequently, coordinate and optimize them in a way that optimal outcomes can be reached.

ProMES incorporates regular, fair, and objective feedback. On the one hand, this serves as a source of information helping employees to adapt their actions in a way that an optimization "from within" can take place. For example, if employees do something differently, they can check in the next performance feedback report if that change was beneficial to the result or not. Beneficial changes can be kept and reinforced, whereas changes that are detrimental to the result can quickly be given up. Moreover, feedback is motivating by allowing for controlling and adjusting of effort, intentions, and alertness by set-actual comparison (Kluger & DeNisi, 1996).

1.4 References

Ashforth, B. E., & Sacks, A. M. (2000). Personal control in organizations: A longitudinal investigation with newcomers. *Human relations, 53,* 311-339.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 191-215.

Bandura, A. (1997). *Self-efficacy. The exercise of control.* New York: Freeman.

Bolfing, A., Halbherr, T., & Schwaninger, A. (2008). How image based factors and human factors contribute to threat detection performance in x-ray aviation security screening. In A. Holzinger (Ed.), *HCI and Usability for Education and Work* (pp. 419-438). Berlin and Heidelberg: Springer.

Brehm, J. W. (1966). *A theory of psychological reactance.* New York, NY: Academic Press.

Brehm, S. S., & Brehm, J. W. (1981). *Psychological reactance: A theory of freedom and control.* New York: Academic Press.

Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research.* Rochester: University of Rochester Press.

Erez, M., Earley, P. C., & Hulin, C. L. (1985). The impact of participation on goal acceptance and performance: A two-step model. *Academy of Management Journal, 28,* 50-66.

European Commission (2006). *Regulation No 1546/2006.* Brussels: European Commission.

Hardmeier, D., Hofer, F., & Schwaninger, A. (2005). The x-ray object recognition test (x-ray ort) – a reliable and valid instrument for measuring visual abilities needed in x-ray screening. In L. D. Sanson (Ed.), *Proceedings of the 39th International Carnahan Conference on Security Technology* (pp. 189-192). Las Palmas: Institute of Electrical and Electronics Engineers.

Harley, B. (1999). The myth of empowerment: Work organization, hierarchy and employee autonomy in contemporary Australian workplaces. *Work, Employment, & Society, 13,* 41-66.

Hofer, F., & Schwaninger, A. (2005). Using threat image projection data for assessing individual screener performance. *WIT Transactions on the Built Environment, 82,* 417-426.

Hofer, F., Wetter, O. E., Graf, R., Guo, Y., & Schwaninger, A. (2009). The use of learning content management systems as a highly adaptive and efficient tool for sustainably supporting (recurrent) training – an applied perspective. In L. D. Sanson & S. M. Steiner-Koller (Eds.), *Proceedings of the 43rd Carnahan Conference on Security Technology* (pp. 52-58). Zürich: Institute of Electrical and Electronics Engineers.

IEEE (2011). *IEEE – the world's largest professional association for the advancement of technology.* Retrieved November 1, 2011, from http://www.ieee.org/index.html.

Klein, H. J., Wesson, M. J., Hollenbeck, J. R., & Alge, B. J. (1999). Goal commitment and the goal-setting process: Conceptual clarification and empirical synthesis. *Journal of Applied Psychology, 84,* 885-896.

Kleingeld, A., van Mierlo, H., & Arends, L. (2011). The effect of goal setting on group performance: A meta-analysis. *Journal of Applied Psychology, 96,* 1289-1304.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119,* 254-284.

Koller, S. M., Drury, C. G., & Schwaninger, A. (2009). Change of search time and non-search time in x-ray baggage screening due to training. *Ergonomics, 52,* 644-656.

Koller, S. M., Hardmeier, D., Michel, S., & Schwaninger, A. (2008). Investigating training, transfer, and viewpoint effects resulting from recurrent CBT of x-ray image interpretation. *Journal of Transportation Security, 1*, 81-106.

Koller, S., & Schwaninger, A. (2006). Assessing x-ray image interpretation competency of airport security screeners. In S. Pejčić-Tarle (Ed.), *Proceedings of the 2nd International Conference on Research in Air Transportation* (pp. 399-402). Belgrade: ICRAT.

Latham, G. P., Erez, M., & Locke, E. A. (1988). Resolving scientific disputes by the joint design of crucial experiments by the antagonists: Application to the Erez–Latham dispute regarding participation in goal setting. *Journal of Applied Psychology, 73,* 753-772.

Latham, G. P., & Locke, E. A. (2007). New developments in and directions for goal-setting research. *European Psychologist, 12,* 290-300.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance.* Englewood Cliffs: Prentice-Hall.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57,* 705-717.

McCarley, J. S., Kramer, A. F., Wickens, Ch. D., Vidoni, E. D., & Boot, W. R. (2004). Visual skills in airport-security screening. *Psychological science, 15,* 302-306.

McClelland, D. C. (1987). *Human motivation.* Cambridge: Cambridge University Press.

Meneer, T., Barrett, D. J. K., Phillips, L., Donnelly, N., & Cave, K. R. (2007). Costs in searching for two targets: Dividing search across target types could improve airport security screening. *Applied cognitive psychology, 21,* 915-932.

Miller, K. I., & Monge, P. R. (1986). Participation, satisfaction, and productivity: A meta-analytic review. *Academy of Management Journal, 29,* 727-753.

Miron, A. M., & Brehm, J. W. (2006) Reactance theory – 40 years later. *Zeitschrift für Sozialpsychologie, 37,* 9-18.

Mitchell, T. R. (1973). Motivation and participation: An integration. *The Academy of Management Journal, 16,* 670-679.

Naylor, J. C., Pritchard, R. D., & Ilgen, D. R. (1980). *A theory of behavior in organizations.* Michigan: Academic Press.

Pritchard, R. D. (1990). *Measuring and improving organizational productivity: A practical guide.* New York: Praeger.

Pritchard, R. D., & Ashwood, E. L. (2008). *Managing motivation: A manager's guide to diagnosing and improving motivation.* New York: Routledge.

Pritchard, R. D., Culbertson, S. S., Malm, K., & Agrell, A. (2009). Improving performance in a Swedish police traffic unit: Results of an intervention. *Journal of Criminal Justice, 37,* 85-97.

Pritchard, R. D., Harrell, M. M., DiazGranados, D., & Guzman, M. J. (2008). The productivity measurement and enhancement system: A meta-analysis. *Journal of Applied Psychology, 93,* 540-567.

Sagie, A., Elizur, D., & Koslowsky, M. (1995). Decision type, participative decision making (PDM), and organizational behavior: An experimental simulation. *Human Performance, 8,* 81-94.

Schuler, R. S. (1980). A role and expectancy model of participation in decision making. *Academy of Management Journal, 23,* 331-340.

Schwaninger, A. (2003). Training of airport security screeners. *Airport, 05/2003,* 11-13.

Schwaninger, A. (2004). Computer based training: A powerful tool to the enhancement of human factors. *Aviation Security International, 02/2004,* 31-36.

Schwaninger, A. (2005). Objekterkennung und Signaldetektion [Object recognition and signal detection]. In B. Kersten (Ed.), *Praxisfelder der Wahrnehmungspsychologie* [Practical fields of cognitive psychology] (pp. 108-132). Bern: Huber.

Schwaninger, A., Hofer, F., & Wetter, O. E. (2007). Adaptive computer-based training increases on the job performance of x-ray screeners. In L. D. Sanson (Ed.), *Proceedings of the 41st Carnahan Conference on Security Technology* (pp. 117-124). Ottawa: Institute of Electrical and Electronics Engineers.

Scully, J. A., Kirkpatrick, S. A., & Locke, E. A. (1995). Locus of knowledge as a determinant of the effects of participation on performance, affect and perceptions. *Organizational Behavior and Human Decision Processes, 61,* 276-288.

Seidenstat, P. (2009). Protecting airline passengers: An overview. In P. Seidenstat & F. X. Splane (Eds.), *Protecting airline passengers in the age of terrorism* (pp. 3-12). Santa Barbara: ABC-CLIO.

Seijts, G. H., & Latham, G. P. (2000). The construct of goal commitment: Measurement and relationships with task performance. In R. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment* (pp. 315–332). Dordrecht: Kluwer Academic.

Shanks, N. E. L., & Bradley, A. L. W. (2004). *Handbook of checked baggage screening: Advanced airport security operation.* London: Professional Engineering Publishing.

Smith, C. S., & Brannick, M. T. (1990). A role and expectancy model of participative decision-making: A replication and theoretical extension. *Journal of Organizational Behavior, 11,* 91-104.

Solak, S., Clarke, J.-P. B., Johnson, E. L. (2009). Airport terminal capacity planning. *Transportation Research, 43,* 659-676.

Srinivasan, N., & Prabhakaran, S. (2009). Current issues in aviation security. In P. Seidenstadt & F. X. Splane (Eds.), *Protecting airline passengers in the age of terrorism* (pp. 247-258). Santa Barbara: ABC-CLIO.

Wales, A. W. J., Anderson, C., Jones, K. L., Schwaninger, A., & Horne, J. A. (2009). Evaluating the two-component inspection model in a simplified luggage search task. *Behavior Research Methods, 41,* 937-943.

Wegge, J. (2004). *Führung von Arbeitsgruppen [Management of teams].* Göttingen: Hogrefe.

Wetter, O. E., Fuhrmann, H., Lipphardt, M., & Hofer, F. (2011). Bringing adversaries together: The importance of a common management-level approach in complex work domains. In M. Faundez-Zanuy, V. Espinosa-Duró, & L. D. Sanson (Eds.), *Proceedings of the 45th Carnahan Conference on Security Technology* (pp. 20-25). Mataró: Institute of Electrical and Electronics Engineers.

Wetter, O. E., Hardmeier, D., & Hofer, F. (2008). Covert Testing at Airports: Exploring Methodology and Results. In L. D. Sanson & K. Fliegel (Eds.), *Proceedings of the 42nd Carnahan Conference on Security Technology* (pp. 357-363). Prague: Institute of Electrical and Electronics Engineers.

Wetter, O. E., Hofer, F., & Jonas, K. (submitted). Efficient and effective goal setting for short interventions in airport security.

Wetter, O. E., Laube, R., & Hofer, F. (2009). A comparison between two leadership models for security checkpoints – enhancing processes by optimizing team performance. In L. D. Sanson & S. M. Steiner-Koller (Eds.), *Proceedings of the 43rd Carnahan Conference on Security Technology* (pp. 59-67). Zürich: Institute of Electrical and Electronics Engineers.

Wetter, O. E., Lipphardt, M., & Hofer, F. (2010). External and internal influences on the security control process at airports. In D. A. Pritchard & L. D. Sanson (Eds.), *Proceedings of the 44th International Carnahan Conference on Security Technology* (pp. 301-309). San José: Institute of Electrical and Electronics Engineers.

Wetter, O. E., Wegge, J., Jonas, K., & Schmidt, K.-H. (submitted). Can dual goals for speed and accuracy on the same performance task prevent speed-accuracy trade-offs?

Widmeyer, W. N., & Ducharme, K. (1997). Team building through team goal setting. *Journal of Applied Sport Psychology, 9,* 97-113.

Winters, D., & Latham, G. P. (1996). The effect of learning versus outcome goals on a simple versus a complex task. *Group and Organization Management, 21,* 236–250.

Wolfe, J. M., Horowitz, T. S., Van Wert, M., J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General, 136,* 623-638.

Wood, R. E., Mento, A. J., & Locke, E. A. (1987). Task complexity as a moderator of goal effects. *Journal of Applied Psychology, 17,* 416–425.

2 Efficient and effective goal setting for short interventions in airport security[1]

Olive E. Wetter, Kantonspolizei Zürich, Research & Development, Zurich-Airport

Franziska Hofer, Kantonspolizei Zürich, Research & Development, Zurich-Airport

Klaus Jonas, University of Zurich, Division of Social and Business Psychology, Zurich

Abstract

Purpose *– This field study investigates the practicability and the effectiveness of minimalistic goal setting prior to short interventions on crew performance. It builds a bridge from goal setting theory to practice and prepares the grounds for its application in security or emergency services.*

Design *– Supervisors of Security Officers at a large European airport communicated goals (no goal, "do your best," "specific difficult goal") without any prior intervention or training to their subordinates. Goals were applied to the "peak hour" only, that is, a short time span of about 40 minutes. Dependent variables were objective measures of efficiency, namely passenger density and throughput, which are true team-level measures, as well as questionnaire measures.*

Findings *– Goal setting can be used easily and effectively by supervisors to increase subordinates' team performance during short interventions. Supervisors could apply goal setting without prior training. Goal setting for short time spans is effective even without providing feedback.*

[1] This study is submitted to the Journal of Business and Psychology.

Implications *– Pre-intervention goal setting is recommended to be installed as a meaningful complement to programs such as "management by objectives", which focus on longer time periods.*

Originality *– This is the first study to investigate effects of goal setting by supervisors without any prior instruction, training, or other interventions. It is equally the first study to analyze effects of short-term goal setting on crew performance in aviation security. In particular, goal setting for a defined short period is studied, allowing for an evaluation of the effectiveness of only the first part of the goal setting process, i.e., without feedback loops.*

2.1 Introduction

Productivity is becoming more and more important in various economic sectors. One main area experiencing huge economic growth is aviation. In just the last five years (2004-2009), commercial aviation in Switzerland has experienced more than 23% increase in the number of departing passengers (Swiss Federal Statistical Office, 2010). One main bottleneck within an airport is the security control process. Because many security checkpoints were planned and constructed based on smaller passenger numbers, the security control process has now become a challenge. Many European airports do not have enough space to physically enlarge the security checkpoints. Therefore, alternative ways have to be found to increase productivity of security controls.

One possible way is the optimization of leadership. Numerous studies demonstrated an impact of leadership on subordinates' performance–for example, in teams and crews (e.g., Guzzo & Dickson, 1996; Zaccaro, Rittman, & Marks, 2001). Optimizing leadership thus promises to be both an effective and cost-saving solution to increase performance of the human workforce. It was decided to open a new focus on productivity of security controls in this study due to the abundance of research that had analyzed performance regarding threat detection only (e.g., Basner et al., 2008; Schwaninger, Hardmeier, Riegelnig, & Martin, 2010; Van Wert, Horowitz, & Wolfe, 2009; Wales, Anderson, Jones, Schwaninger, & Horne, 2009). Moreover, Wetter, Fuhrmann, Lipphardt, and Hofer (2011) stated that the objectives of the security control at the respective airport, as defined by the management, contain two most important indicators that are of equal importance.

Leaders can set goals for their subordinates and thus influence the way they work and perform. In organizations, "management by objectives" (e.g., Drucker, 1974, 2007; Kopelman, 1986; Odiorne, 1978) has become popular. Following this method, goals are set for a predefined specific time period, which is usually 6-12 months. During this period, goals are in place, the employee works with them, and data are collected. After this time period, goal progress is evaluated together with the supervisor, and, subsequently, a new goal is set for the next time period. In the present study, we argue that setting goals for short time periods (e.g., for an intervention that lasts 1 hr) can be a way of positively influencing subordinates' work performance that is both effective and cost-efficient. Intelligent goal setting thus presents itself as one of the ideal solutions for optimizing or increasing performance during short interventions that are common for example in emergency services.

Several aspects are inherent to goals, such as goal difficulty or specificity. Goal setting theory by Locke and Latham (1990, 2002) outlines different aspects that characterize effective goals. One of the theory's most important tenets is that specific, difficult goals lead to better performance than generalized "do your best" goals (Locke & Latham, 1990). Already in the 1970s there were sufficient field studies available for Latham and Yukl (1975) to review. Today, a large body of research documents different aspects and effects of goal setting in the laboratory and the field. Locke and Latham (2006) summarized some of the most important newer developments, some of which are group goal setting (e.g., Haslam, Wegge, & Postmes, 2009), nonconscious goals (e.g., Eitam, Hassin, & Schul, 2008), and applications in other disciplines such as medicine (e.g., Bradley, Bogardus Jr., Tinetti, & Inouye, 1999) or sports (e.g., Kingston & Wilson, 2009).

We would like to suggest that research on goal setting can be divided into two categories: First, there are laboratory studies that typically examined the effects of setting a goal that is valid for a short time period only, as laboratory experiments typically last less than one hour. This kind of goal setting will subsequently be called "short-term". Short-term goal setting in laboratory studies typically involved time frames of minutes to hours. Examples of studies using a short-term time span are Locke (1982), with a time frame of 1 min, and Terborg and Miller (1978), with a time frame of 2 hr. Second, there are applied studies, in which goals were set and in place during periods of time that were much longer. This kind of goal setting will subsequently be called “long-term”. Examples of applied studies using this long-term time span range from Latham and Kinne (1974), where goals remained in effect at least for the existing day, to Ivancevich and McMahon (1982) or Buller and Bell Jr. (1986), where a goal was set and remained valid for a period of three months.

This is where the present study ties in and tries to close three gaps. First, it examines the effects of short-term goal setting in the field by explicitly setting a goal that is valid for a defined short period only (about 40 min). Our design allows for an evaluation of the effectiveness of only the first part of the common goal setting process, i.e., goal setting and first performance by the employee, without feedback and subsequent adjustment of behavior by the employee or even multiple feedback loops. Second, to our knowledge, there is as yet no field study that examined the effects of minimalistic goal setting by supervisors, meaning without any instruction, training, or other interventions that cost time and money. Third, this is the first study to analyze effects of short-term goal setting on crew performance and way of working in aviation security. To this end, crew performance in terms of efficiency was measured using two true crew-level measures, namely, passenger density and throughput in security control lanes. Further, aspects of employees’ cognitive representations of goals and their effects are analyzed and discussed.

2.2 Method

2.2.A Setting and Participants

This study was conducted at the security control of a large European airport that serves as an airline hub. Data were collected from early February to mid March 2010 (six weeks). To ensure highest possible standardization, data collection took place always on the same days (Wednesday, Thursday, and Friday) during the morning peak hour between 6 a.m. and 7 a.m. A total of six security control screening lanes, i.e., all lanes for passengers with class code “Y” (economy) in the most important sector at the airport, were studied. The lanes all featured the same equipment, including a Smiths Heimann Hi-Scan 6046si X-ray device and a Ceia 02PN20 walk-through metal detector (WTMD). The layout of the six lanes was very similar but not identical due to constructional constraints.

The security control process basically requires the passenger to put baggage and belongings on the roller table, pass the WTMD, and collect baggage and belongings again after X-ray inspection. The security control is carried out by Security Officers (SOs) between 20 and 65 years of age, that are employed by a governmental organization. Although SOs are not police officers, the working environment is embedded in police structures and is thus in some respects comparable to police work. All SOs and supervisors are trained and certified according to the actual standards by the country's regulatory authority. SOs have no fixed shift schedule, so crews consist of five different employees every day[2]. Crew composition is ad hoc, changes on a daily basis, and is random with the exception that some criteria have to be followed (e.g., both sexes have to be represented in a crew). Allocation of crews to the studied sector was fully random. As a consequence, crews were randomly assigned to the three different experimental conditions (see paragraph on variables below).

Each crew operates one screening lane and consists of four SOs and one crew leader, who is in charge of the crew. In the sector of the airport where this study was conducted, two supervisors are responsible for controlling and problem solving and have higher decision making authority. One of them is responsible for holding pre-shift briefings. This study was conducted during standard daily operation in order to ensure maximal ecological validity. Accordingly, we did not change the shift plans. Due to that, some SOs were present at more than one condition in the study, but crews were never the same at different measurements.

---

[2] One day during the measurements, there was a shortage of personnel. Due to that, a crew was formed consisting of 3 SOs plus a supervisor, who performed the duties of an SO.

2.2.B Variables

The study featured one independent variable, namely, goal setting, with three different conditions: one baseline and two experimental conditions, which we labeled "do your best" and "specific difficult goal." In the baseline condition, no intervention was made and no goal communicated to the SOs. This condition reflects standard daily operation with the only exception that two observers were present behind the six screening lanes (which was the same in all three conditions). The "do your best" condition used a classical "do your best" goal as the intervention. Prior to the shift, staff from the organization's research & development (R&D) unit contacted the supervisor in charge of the pre-shift briefing for the crews that were studied subsequently. The supervisor was given a memo stating the following: "This morning it is our goal that everybody does their best in order to achieve excellent job performance" (here freely translated from the German). The supervisor was asked to communicate this information to the crews at the end of the pre-shift briefing.

The "specific difficult goal" condition used a specific, difficult goal as the intervention. The supervisor was given a memo stating the following: "In today's morning peak, in addition to the compliance with security regulations as usual, it is our goal to ensure that there are always 8 passengers simultaneously in a screening lane. We, as members of the present shift, can occasionally monitor that ourselves" (freely translated here). In other words, crews were told that they can monitor their own goal attainment to a certain extent, if they like. Again, the supervisor was asked to communicate this instruction to the crews at the end of the pre-shift briefing. This goal was chosen because passenger density is important (see paragraph on passenger density below), because it is specific (it contains a number), and because it is known to be difficult but realistic based on a preliminary study by Wetter, Lipphardt, and Hofer (2010). This preliminary study demonstrated that, given this particular airport's infrastructure, this number is usually lower (around 6), but can reach 8 passengers if crews are motivated and work together well. To avoid spillover effects from the specific condition to the more unspecific conditions, it was decided to first carry out all baseline measurements, then all measurements with the "do your best" goal, and, finally, all measurements with the specific difficult goal; all within six weeks.

Crew performance was assessed in terms of efficiency. Two important efficiency measures were recorded, namely passenger density and throughput. Passenger density is a measure of the degree of infrastructure capacity utilization. It is receiving more and more attention in the applied field, since it denotes how well the space of a given security control infrastructure is actually used. As every m^2 in an airport is very costly, security control would ideally be limited to as many m^2 as necessary, with as little "empty" or unused area inside as possible. In a given infrastructure, higher passenger density indicates that less space is occupied per passenger, i.e., that the available space is better utilized. Further, this measure is interesting, because it can be controlled by the SOs themselves via the way they work and guide passengers. Lower than optimal passenger density would be considered as "wasting space" and economically inefficient, whereas higher than optimal passenger density could lead to disorderly, chaotic situations, which in turn could lower the security level, increase employees' work strain, and negatively affect passenger satisfaction.

Passenger density can be defined as the number of passengers present simultaneously in one screening lane at one moment. An important precondition for measuring passenger density is that there are enough passengers queuing up. That is, at least one passenger has to be in a waiting position. The number of passengers present simultaneously is calculated as follows: The first passenger to be counted is the one who is at the very beginning of the security control process, i.e., the one who does the first action (usually touching the baggage tray). The last one to be counted is the one who is at the very end of the security control process (usually the one stepping away from the control lane). All passengers within this process, for example passengers loading their hand baggage into the tray, passing through the WTMD, undergoing a pat-down search, or collecting their belongings, are counted.

To ensure a certain minimum distance between measurements, a maximal sampling frequency of 1/min was defined. Sampling was done visually by two observers who were present behind the security control screening lanes. Each of those observers sampled data from three security control lanes sequentially, i.e., started at lane one and sampled, then moved on to lane two and sampled, then moved on to lane three and sampled, then began again with lane one. Like this, a certain number of samplings of passenger density per control lane resulted after the measurement period. Subsequently, those were averaged into one value of passenger density per control lane per measurement period.

The second efficiency measure was passenger throughput. Throughput is very important for different stakeholders, because it is used for capacity calculations and for infrastructure and personnel planning. It is also important for ensuring certain minimum connecting times for transferring passengers and for the calculation of queuing times for security control. If throughput is neglected, “airlines or airports whose security screening procedures are particularly time consuming can expect to lose business to their competitors” (Gilliam, 1979, p. 117). Throughput is a widely recognized measure in the applied field and is used for benchmarking between European airports. We propose the following definition and operationalization of passenger throughput: Passenger throughput denotes the maximum capacity of a security control lane in passengers/hr.

As it is a measure denoting maximum capacity, it must be ensured that during measurement there are more passengers available than the number that the security control lane can actually handle. If the number of passengers is lower than the capacity of the security control lane, the measure only denotes the number of passengers and is not an indicator of maximal capacity. Therefore, measurements must only take place if there is a queue of passengers, i.e., at least one passenger queuing up and waiting in front of the lane. Throughput was measured automatically using the relevant function of the WTMD's standard commercial software. As demonstrated in the preliminary study (Wetter et al., 2010), the absolute values of data collected automatically by the WTMDs in the respective sector at the respective airport are too high by approximately M = 15.1%, SD = 7.3%[3]. However, further analyses by the same authors proved that automatically collected data can be used very well for analyzing variations over time at the same location. This is exactly what they were used for in this study, where absolute values are of little importance.

Goal setting is expected to affect passenger density because it can be controlled by the SOs themselves. However, it is unclear if it will also affect throughput. Apart from employee behavior, throughput significantly depends on external factors as well, such as "number of manual baggage inspections to be done" and "day temperature" (Wetter et al., 2010). Throughput was chosen as second measure for efficiency because it is known to correlate with passenger density (Wetter et al., 2010) and because it is widely used. Both are true crew-level measures – that is, one value per crew is generated. Further, the two dependent variables were collected for all crews present (population level) in the respective sector during the respective time interval.

In addition to these measured variables, employee self-report measures were collected with a questionnaire, which mainly served for assessing goal recall performance after the short intervention. The employees were asked the question, "Evaluation of the briefing: What goal were you assigned by the supervisor at today's briefing?" Possible answers were "There was no goal assigned" (with a check box) or "The following goal was assigned:" (with an empty line for writing in the goal). In this way, it was assessed whether SOs thought a goal had been assigned, and if yes, what

[3] For instance, SOs passing the WTMD or passengers passing the WTMD twice contribute to an overestimation of absolute values for throughput in automatic measurements.

they thought the goal had been. In the baseline condition, the questionnaire also contained items to assess the representativeness of the work conditions during the study as compared to everyday work conditions regarding throughput, security level, and friendliness towards passengers. Seven questions about subjective impressions of different work aspects were asked in both the baseline and the “do your best” condition to compare if the impressions changed along with goal setting. In addition, there were a few items concerning the employees’ views on goal attainment in the “do your best” condition only. For operational reasons, it was decided to skip this in the “specific difficult goal” condition and to have the SOs answer the goal recall question only.

2.2.C Procedure

Prior to the work shift, the supervisor in charge of the pre-shift briefing in the sector was told that staff persons from R&D would be present during the morning peak hour. For the “do your best” and the “specific difficult goal” condition, the supervisor was given a memo with the goal (see section “Variables”). The supervisor then communicated this information to the SOs at the briefing. Some supervisors communicated the goal by simply reading out the memo aloud; others communicated the goal using their own words. This communication process was not fully standardized because it is known that this is exactly what happens in everyday practice: Every supervisor communicates new instructions in his own way. Thus, the study was devised to provide maximal ecological validity. During the briefing, staff persons from R&D were present as passive listeners. After the briefing, SOs opened the assigned security control lanes to passengers. As soon as there were enough passengers waiting, that is, when there were queues and no more gaps between passengers, measurements started. In order to measure passenger density, two people from R&D circulated behind the security control lanes, each of them observing three lanes. The measurement period for both, passenger density and throughput, stopped as soon as there were gaps between passengers approaching the security control lanes instead of continuous queues. Measurements lasted 39 min on average (*M* = 39.34 min, *SD* = 9.42 min). A few minutes afterwards, the questionnaire was administered to the SOs.

The baseline condition was devised in order to guarantee the same amount of observer presence and for recording and measuring passenger density in exactly the same way as in the conditions with goal setting. Thus, the procedure of the baseline condition was identical to the one described above with the only difference that no memo was given to the supervisor, and, as a consequence, no goal setting took place.

2.3 Results

2.3.A Manipulation Check

The observers always noted during the supervisor's briefing if the instruction from the memo was in fact communicated to the crews, which served as manipulation check. This was always the case, thus, the success rate of the manipulation check being 100%.

2.3.B Interrater Reliability for Passenger Density

Interrater reliability for passenger density was established as follows: After measurements were finished, both observers[4] plus a third person who had not been involved in data collection measured the same security control lanes in an additional session. Thirty measurements were recorded independently by these three observers, and they were subsequently compared applying Pearson correlations. Ratings of the two "trained" observers who collected data for the study correlated with r = .94. The correlation between the ratings of observer 1 and the "new" observer was r = .92; the correlation between observer 2 and the "new" observer was r = .91.

[4] The first author was one of the observers.

2.3.C Efficiency

Because it makes sense to integrate both, passenger density and throughput, into a construct "efficiency", and because the preliminary study (Wetter et al., 2010) had demonstrated that they are correlated with r = .27, a MANOVA was calculated. The factor (between) was "Goal" (baseline, "do your best," "specific difficult goal"). In the multivariate analysis, the dependent variable was "efficiency" as a construct. The MANOVA revealed a significant effect of "Goal" with $F(4, 196) = 5.006$, $p < .001$ and an effect size of $\eta_p^2 = .093$ on efficiency. In order to find out which of the two dependent variables contributed to this result, univariate analyses were calculated.

Univariate tests revealed that "Goal" had a significant effect on passenger density with $F(2, 98) = 10.971$, $p < .001$ and an effect size of $\eta_p^2 = .183$. Simple contrasts revealed significant differences between all possible comparisons: baseline vs. "do your best" with $p < .05$, "do your best" vs. "specific difficult goal" with $p < .05$, baseline vs. "specific difficult goal" with $p < .001$ (Figure 1, left). On the other hand, there was no significant effect of "Goal" on throughput with $F(2, 98) = 0.177$, $p = .838$ (Figure 1, right).

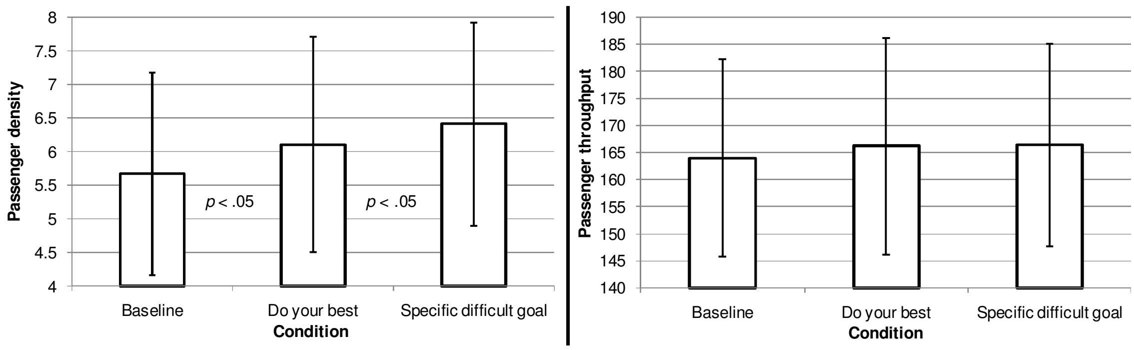

Figure 1. Means and standard deviations for passenger density (left) and throughput (right) for the three conditions.

2.3.D Goal Recall

The questionnaire containing the goal recall question was filled out individually by more than 79% of the SOs overall. Figure 2 (see next page) shows the exact numbers for the three conditions. In the baseline condition, the question ("What goal was assigned by the supervisor at today's briefing?") revealed that SOs had no goals in mind: Out of a total of $N = 96$ individual responses, 56% cited a new regulation that had nothing to do with a goal, 24% cited other information that had been communicated in the briefing, 19% stated explicitly that no goal had been communicated, and 1% reported "do not know."

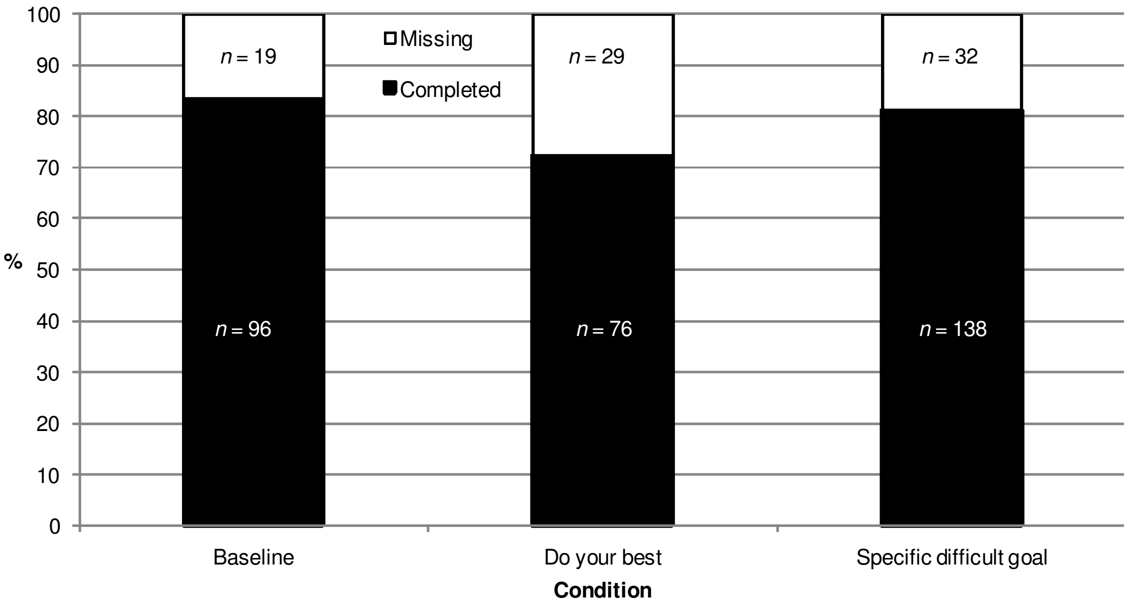

Figure 2. Questionnaire completed and missing for the three conditions.

Since the dependent variables are crew-level measures, answers to the goal recall question for the "do your best" and the "specific difficult goal" condition were categorized as either "correct" or "wrong" for the whole crew. They were regarded as "correct" if at least three out of five crew members stated the correct goal. Otherwise, they were regarded as "wrong" for the whole crew. On the individual level, correctness of the reported goal was judged conservatively[5]. Figure 3 (see next page) shows the goal recall performance in the "do your best" and the "specific difficult goal" conditions.

[5] For example in the "specific difficult goal" condition, the answers "max. 8 passengers" or "work fast to ensure good passenger flow" were counted as wrong. The first one is conceptually wrong, and the second one has lost its specific, difficult character.

Free recall of the goal succeeded significantly more often in the “specific difficult goal” condition than in the “do your best” condition, with $\chi^2(1, N = 137) = 26.12$, $p < .001$.

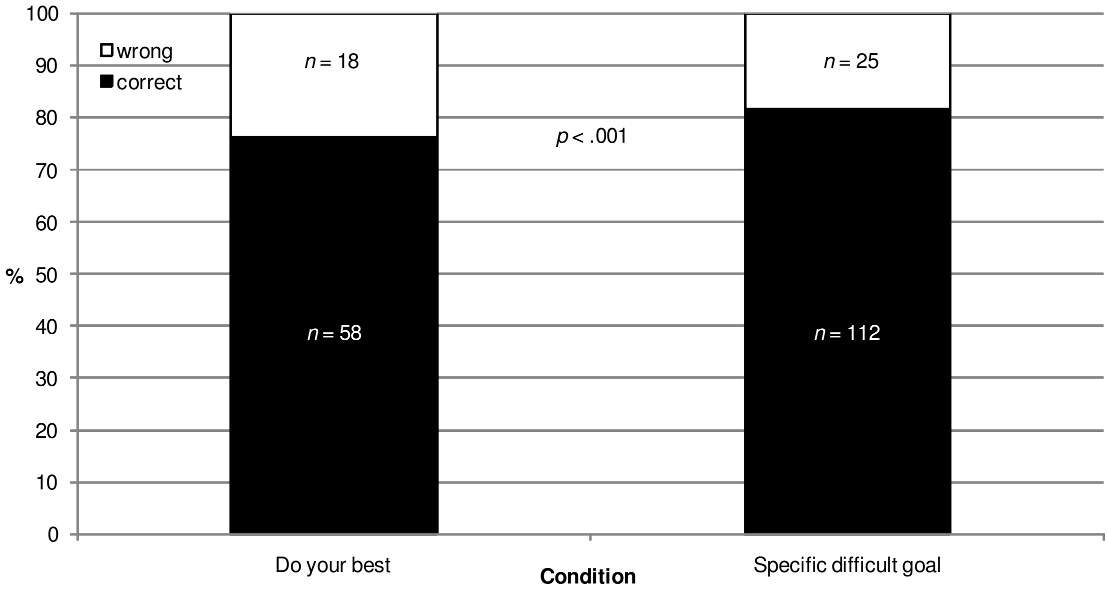

Figure 3. Goal recall performance for the “do your best” and “specific difficult goal” condition.

2.3.E Subjective Impressions of SOs

While working in the field, it was discovered that the response format of two questionnaire items was wrong and misleading[6]. After discovery, this was corrected for the subsequent day of data collection, and data from questionnaires with the misleading format were excluded. Somewhat later, when analyzing questionnaire data, it was discovered that two other items were poorly phrased. One item was obsolete, and the other one did not make sense[7]. Due to that, it was decided not to analyze

[6] The items “as compared to an average hour of work, the work strain was” and “as compared to an average hour of work, my tiredness was” first used the response format “much worse – same – much better,” which could not be interpreted and does not make sense.

[7] The questions, “How clear was the goal assigned by the supervisor?” and “Do you think you neglected other aspects of work due to goal setting at the briefing?” were not analyzed. The first one is obsolete, because this is tested with the goal recall. The second one does not make sense for the “do your best” condition, because this goal is completely unspecific and does not focus on any aspect.

those two items at all. After these exclusions, the SOs' ratings from 1 to 5 were analyzed (Figure 4).

Subjective ratings reached average values of $M = 3.04$ for throughput, $M = 3.14$ for security, and $M = 3.20$ for friendliness towards passengers. As a rating of 3 represents exactly the middle in our response format, it can be said that baseline ratings are very close to the middle. This indicates that during our baseline measurements, SOs did not perceive any of the subjective variables to be extraordinarily high or low. As we wanted to compare conditions with goal setting to the current practice without goal setting, this is an important result in order to demonstrate that the baseline condition stands representatively for standard daily operation.

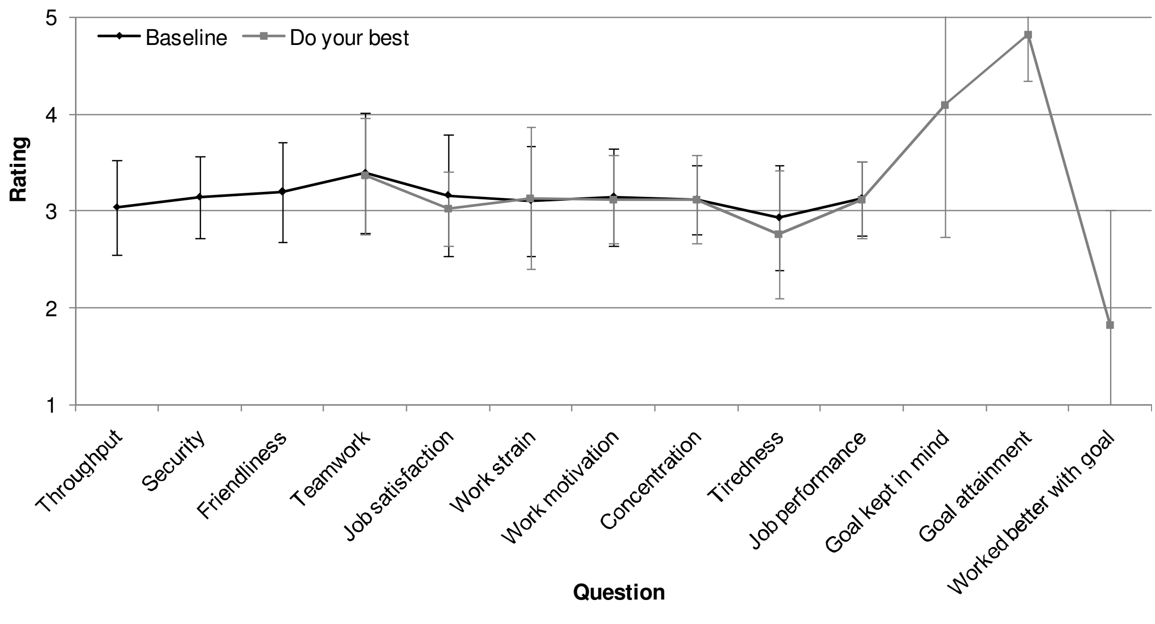

Figure 4. Ratings of questionnaire (means and standard deviations) for the baseline and "do your best" conditions: the higher the rating, the "higher" the subjective impression in the respective aspect. First three questions in baseline condition only, last three questions in "do your best" condition only.

Seven items were answered in both the baseline and the "do your best" condition. Ratings for these seven items were compared using two-tailed *t* tests. For these pairwise comparisons, a Bonferroni correction (e.g., Abdi, 2007; Bortz, 1999) was applied. Results from all comparisons lack significance (all *p*s > .060). These results show that subjective impressions of several work aspects did not differ significantly between the baseline and the "do your best" condition. The three items that had only been asked in the "do your best" condition revealed one interesting result: Although SOs were of the impression that they had kept the goal in mind reasonably well during work (M = 4.09, SD = 1.35), and although they thought they attained it quite well (M = 4.82, SD = 0.47), they did not think they worked better with the goal (M = 1.83, SD = 1.18).

2.4 Discussion

The results demonstrate that goals for a time span of about 40 min can significantly increase crew performance of SOs in terms of efficiency at work. This increase in efficiency came about even though there was no feedback and subsequent possibility for adjusting behavior. Moreover, the increase in performance was caused by minimalistic goal setting by supervisors, i.e., without prior instruction, training, or other interventions that cost time or money. It could be replicated that specific, difficult goals lead to better performance in the targeted specific aspect than an unspecific "do your best" goal. With regard to memory, the specific, difficult goal could be better recalled after about 40 min than the "do your best" goal. Contrary to our expectations, employees themselves did not perceive their increase in performance.

In many studies, "do your best" was used as baseline condition (e.g., Latham & Baldes, 1975; Locke & Latham, 1990). In the present study, "do your best" could be compared to no goal at all, which served as the baseline condition. This was possible, because SOs know their task well and work according to their directives. Because it lacks an external referent, the "do your best" goal was defined idiosyncratically (Locke & Latham, 2002) by the SOs. When SOs tried to "do their best," passenger density in the security control lanes was significantly higher than the baseline performance without goal setting. This effect of "do your best" is noteworthy because it did not make

SOs focus on any specific aspect of their work. It may therefore be speculated that an unspecific “do your best” goal has also beneficial effects on many other aspects and variables apart from passenger density. Focusing SOs on passenger density alone by using a specific, difficult goal, led to even better performance in this specific aspect. This finding of an advantage of a specific, difficult goal as compared to a “do your best” goal with regard to the variable addressed is fully in line with earlier research (Locke & Latham, 1990, 2002).

The preliminary study (Wetter et al., 2010) revealed a positive correlation of $r = .27$ between passenger density and throughput. However, that finding was based on a small sample size. With the present study, the finding could be replicated ($r = .23$, $p < .05$) with $N = 101$. Both results provide strong evidence that passenger density and throughput are indeed positively correlated measures of security control efficiency. Thus, handling more passengers simultaneously is beneficial to the overall passenger flow. We learned from personal communication with SOs that with increasing passenger density, perceived stress increases and perceived security decreases. Data from the questionnaire, however, hinted at the fact that if SOs do not consciously know that passenger density is high, perceived work strain does not change. Of course, the absolute number of passenger density is influenced not only by crew performance but also by other factors. One of these is the size of the infrastructure or space in general. Therefore, the maximal value of the variable passenger density is limited depending on infrastructure. It can be speculated that there is an optimal passenger density for each layout or size of infrastructure.

The study revealed no significant effect of goal setting on throughput. There are three possible reasons for this lack of significance. First, SOs were told or instructed to pay attention to throughput in neither the “do your best” nor in the “specific difficult goal” condition. However, they were instructed in the “specific difficult goal” condition to care for passenger density. Second, data from throughput have rather large standard deviations, which make it more difficult to find significant effects. Third, throughput is significantly influenced by external factors, on which crew members have no influence, which is a major difference to passenger density. This limitation can be illustrated by the fact that the two external factors "number of manual baggage inspections to be done[8]" and "day temperature" account for 25% and 22% respectively of the variations in throughput (Wetter et al., 2010).

Interrater reliability for measurements of passenger density between the two observers was very high. Moreover, measurements with a "new" observer reached almost the same reliability values. These results show that measuring passenger density using the proposed method is reliable, can be learned quickly, and is little prone to distortions. The concept is also very simple in practice and leaves little room for interpretation.

Data from the goal recall question revealed that the specific, difficult goal could more easily be recalled after the morning peak hour than the "do your best goal." We suggest that a goal with a specific number can be more easily retained in memory and be subsequently recalled than an unspecific goal. Findings from cognitive psychology strongly support this idea: Walker and Hulme (1999) showed that concrete words are easier to recall than abstract words. Subjective ratings revealed mean values for throughput, security, and friendliness that were very close to 3 in the baseline condition. Therefore, baseline measurements were very close to the average work situation and, thus, representative. The conclusion can be drawn that during baseline measurements, there were no extraordinary events that affected the domains in question, which is necessary for generalization of the findings of this study.

Other ratings showed that SOs thought they did not work better with the specific, difficult goal. However, objective data clearly point in the opposite direction.

---

[8] For one, the number of manual baggage inspections depends on what kind of items passengers carry with them. For another, it is positively influenced by training (Koller, Hardmeier, Michel, & Schwaninger, 2008) and the experience (Schwaninger, Hardmeier, & Hofer, 2005) of the X-ray screener.

This hints at a discrepancy between the subjective evaluation of the goal's effects and the objective results if there is no feedback. In a similar vein, it was found that seven ratings about different work aspects did not differ significantly between the baseline and the "do your best" condition. This might indicate that indeed both work situations were comparable, or – as is more likely if compared with objective measures (significantly higher passenger density) – changes in behavior brought about by goal setting were not consciously represented in SOs' minds. It might be speculated that people underestimate goals and their effects on behavior.

A rival hypothesis that might come to mind is that motivational effects caused by knowledge of results could have caused the increases in performance. Knowledge of results means that learning is facilitated, if participants are informed about the progress they are making. In other words, increases in performance might be due to effects of (continuous) feedback rather than to goal setting. Translated to our applied situation, this could for instance mean that if SOs counted passenger density every now and then, this activity alone motivated them to reach higher and higher numbers. Looking at our design, effects of knowledge of results can be completely ruled out for the "do your best" condition, since it is unspecific. SOs neither knew what to focus on nor did they know which variables were measured by the observers. For the "specific difficult goal" SOs could observe the number of passengers present simultaneously and therefore could gain some knowledge of results. We cannot estimate how often they had time to do so and actually did it. Thus, the alternative explanation that knowledge of results brought about the increase between the do your best and the specific difficult condition cannot be ruled out with our design alone.

Confronted with the same question, Locke (1967) and Locke, Cartledge, and Koeppel (1968) separated effects of knowledge of results and goal setting. They showed that effects are due to different levels of motivation produced by the different goals and not by knowledge of results. It was further shown that knowledge of results does not appear to affect performance independently of the goals set (Dossett, Latham, & Mitchell, 1979). In a similar vein, Locke et al. (1968) stated that the mere presence of knowledge of results does not increase performance unless it is used by the individual to set a specific hard goal. Erez (1977) showed that feedback actually boosts the relationship between goal setting and performance by facilitating the display of individual differences in goal setting.

But what about social facilitation (Zajonc, 1965) or the “Hawthorne Effect” (Landsberger, 1958; Roethlisberger & Dickson, 1939)? In order to prevent those from coming into effect in our study, it was ensured that there was exactly the same amount of presence of supervisors and observers in all conditions including the baseline. The drawback of this is that possible overall reactivity, caused by observer presence, cannot be ruled out. However, the presence of observers was necessary in order to measure passenger density.

Locke, Latham, and Erez (1988, p. 23) stated, “if there is no commitment to goals, then goal setting does not work.” Similar to that, Latham and Yukl (1975, p. 824) stated, “goals that are assigned to a person (e.g., by a supervisor) have an effect on behavior only to the degree that they are consciously accepted by the person.” French, Kay, and Meyer (1966) pointed out, however, that this relation was only observed when employees had a past history of participation. In our study, we used an authoritative way of setting goals without participation by supervisors and employees. Regarding police and security work, where authoritative goal setting is predominant, it can be argued that our setting was ecologically valid. Moreover, Locke and Latham (1990) found a correlation of r = .58 between set goals and personal goals. This suggests that a significant part of authoritatively set goals will later be accepted and treated as personal goals.

The present study demonstrated that minimalistic goal setting can be used at briefings in order to enhance work performance during short interventions. It is easy to use, efficient and effective, provided the tenets of goal setting theory are followed. It can thus be recommended to install pre-intervention goal setting as a meaningful complement to programs such as “management by objectives”, which focus on longer time periods.

2.5 References

Abdi, H. (2007). Bonferroni and Sidak corrections for multiple comparisons. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 103-107). Thousand Oaks, CA: Sage.

Basner, M., Rubinstein, J., Fomberstein, K. M., Coble, M. C., Ecker, A., Avinash, D., & Dinges, D. F. (2008). Effects of night work, sleep loss, and time on task on simulated threat detection performance. *Sleep, 31,* 1251-1259.

Bortz, J. (1999). *Statistik für Sozialwissenschaftler* [Statistics for social scientists]. Berlin, Germany: Springer.

Bradley, E. H., Bogardus Jr., S. T., Tinetti, M. E., & Inouye, S. K. (1999). Goal-setting in clinical medicine. *Social Science & Medicine, 49,* 267-278.

Buller, P. F., & Bell Jr., C. H. (1986). Effects of team building and goal setting on productivity: A field experiment. *The Academy of Management Journal, 29,* 305-328.

Dossett, D. L., Latham, G. P., & Mitchell, T. R. (1979). Effects of assigned versus participatively set goals, knowledge of results, and individual differences on employee behavior when goal difficulty is held constant. *Journal of Applied Psychology, 64,* 291-298.

Drucker, P. (1974). *Management: Tasks, responsibilities, practices.* Oxford, United Kingdom: Butterworth-Heinemann.

Drucker, P. (2007). *The practice of management (Classic Drucker Collection edition).* Oxford, United Kingdom: Butterworth-Heinemann.

Eitam, B., Hassin, R. R., & Schul, Y. (2008). Nonconscious goal pursuit in novel environments: The case of implicit learning. *Psychological Science, 19,* 261-267.

Erez, M. (1977). Feedback: A necessary condition for the goal setting-performance relationship. *Journal of Applied Psychology, 62,* 624-627.

French, J. R., Kay, E., & Meyer, H. H (1966). Participation and the appraisal system. *Human Relations, 19,* 3-19.

Gilliam, R. R. (1979). An application of queuing theory to airport passenger security screening. *Interfaces, 9,* 117-122.

Guzzo, R. A., & Dickson, M. W. (1996). Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology, 47,* 307-338.

Haslam, S. A., Wegge, J., & Postmes, T. (2009). Are we on a learning curve or a treadmill? The benefits of participative group goal setting become apparent as tasks become increasingly challenging over time. *European Journal of Social Psychology, 39,* 430-446.

Ivancevich, J. M., & McMahon, J. T. (1982). The effects of goal setting, external feedback, and self-generated feedback on outcome variables: A field experiment. *The Academy of Management Journal, 25,* 359-372.

Kingston, K. M., & Wilson, K. M. (2009). The application of goal setting in sport. In S. D. Mellalieu, & S. Hanton (Eds.), *Advances in applied sport psychology* (pp. 75-123). New York, NY: Routledge.

Koller, S. M., Hardmeier, D., Michel, S., & Schwaninger, A. (2008). Investigating training, transfer and viewpoint effects resulting from recurrent CBT of x-ray image interpretation. *Journal of Transportation Security, 1,* 81-106.

Kopelman, R. E. (1986). *Managing productivity in organizations.* New York, NY: McGraw-Hill.

Landsberger, H.A. (1958). *Hawthorne revisited: Management and the worker: Its critics, and developments in human relations in industry.* Ithaca, NY: Cornell University.

Latham, G. P., & Baldes, J. J. (1975). The "practical significance" of Locke's theory of goal setting. *Journal of Applied Psychology, 60,* 122-124.

Latham, G. P., & Kinne, S. B. (1974). Improving job performance through training in goal setting. *Journal of Applied Psychology, 59,* 187-191.

Latham, G. P., & Yukl, G. A. (1975). A review of research on the application of goal setting in organizations. *The Academy of Management Journal, 18,* 824-845.

Locke, E. A. (1967). Motivational effects of knowledge of results: Knowledge or goal setting? *Journal of Applied Psychology, 51, 324-329.*

Locke, E. A. (1982). Relation of goal level to performance with a short work period and multiple goal levels. *Journal of Applied Psychology, 67,* 512-514.

Locke, E. A., Cartledge, N., & Koeppel, J. (1968). Motivational effects of knowledge of results: A goal-setting phenomenon? *Psychological Bulletin, 70,* 474-485.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance.* Englewood Cliffs, NJ: Prentice Hall.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. *American Psychologist, 57,* 705-717.

Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science, 15,* 265-268.

Locke, E. A., Latham, G. P., & Erez, M. (1988). The determinants of goal commitment. *Academy of Management Review, 13,* 23-39.

Odiorne, G. S. (1978). MBO: A backward glance. *Business Horizons,* October 1978, 14-24.

Roethlisberger, F. J., & Dickson, J. (1939). *Management and the worker.* Cambridge, MA: Harvard University Press.

Schwaninger, A., Hardmeier, D., & Hofer, F. (2005). Aviation security screeners visual ability and visual knowledge management. *IEEE Aerospace and Electronic Systems Magazine, 20,* 29-35.

Schwaninger, A., Hardmeier, D., Riegelnig, J., & Martin, M. (2010). Use it and still lose it? *GeroPsych, 23,* 169-175.

Swiss Federal Statistical Office (2010). *Mobilität und Verkehr 2010* [Mobility and traffic 2010]. Neuchâtel, Switzerland: Swiss Federal Statistical Office.

Terborg, J. R., & Miller, H. E. (1978). Motivation, behavior, and performance: A closer examination of goal setting and monetary incentives. *Journal of Applied Psychology, 63,* 29-39.

Van Wert, M. J., Horowitz, T. S., & Wolfe, J. M. (2009). Even in correctable search, some types of rare targets are frequently missed. *Attention, Perception, & Psychophysics, 71,* 541-553.

Wales, A. W. J., Anderson, C., Jones, K. L., Schwaninger, A., & Horne, J.A. (2009). Evaluating the two-component inspection model in a simplified luggage search task. *Behavior Research Methods, 41,* 937-943.Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology, 25,* 1256-1271.

Wetter, O. E., Fuhrmann, H., Lipphardt, M., & Hofer, F. (2011). Bringing adversaries together: The importance of a common management-level approach in complex work domains. In M. S. Faundez-Zanuy, V. Espinosa-Duró, & L. D. Sanson (Eds.), *Proceedings of the*

45th International Carnahan Conference on Security Technology (pp. 20-25). Mataró: Institute of Electrical and Electronics Engineers.

Wetter, O. E., Lipphardt, M., & Hofer, F. (2010). External and internal influences on the security control process at airports. In D. A. Pritchard, & L. D. Sanson (Eds.), *Proceedings of the 44th International Carnahan Conference on Security Technology* (pp. 301-309). San José, CA: Institute of Electrical and Electronics Engineers.

Zaccaro, S. J., Rittman, A. L., & Marks, M. A. (2001). Team leadership. *Leadership Quarterly, 12,* 451-483.

Zajonc, R. B. (1965). Social facilitation. *Science, 149,* 269-274.

3 Can dual goals for speed and accuracy on the same performance task prevent speed-accuracy trade-offs?[1]

Olive E. Wetter, Kantonspolizei Zürich, Research & Development, Zurich-Airport

Jürgen Wegge, Technische Universität Dresden, Work and Organizational Psychology, Dresden

Klaus Jonas, University of Zurich, Division of Social and Business Psychology, Zurich

Klaus-Helmut Schmidt, Leibniz Research Centre for Working Environment and Human Factors, Dortmund

Abstract – In most work contexts, several performance goals coexist, and conflicts between them and trade-offs can occur. Our paper is the first to contrast a dual goal for speed and accuracy with a single goal for speed on the same task. The Sternberg paradigm (Experiment I, n = 57) and the d2 test (Experiment II, n = 19) were used as performance tasks. Speed measures and errors revealed in both experiments that dual as well as single goals increase performance by enhancing memory scanning. However, the single speed goal triggered a speed-accuracy trade-off, favoring speed over accuracy, whereas this was not the case with the dual goal. In difficult trials, dual goals slowed down scanning processes again so that errors could be prevented. This new finding is particularly relevant for security domains, where both aspects have to be managed simultaneously.

---

[1] This study is submitted to the Journal of Personnel Psychology.

3.1 Introduction

In many laboratory and field studies, goal setting proved to be an effective and efficient tool to enhance performance of the workforce (Locke & Latham, 2002). Provided that the preconditions for goal setting are favorable (e.g., moderate task complexity, commitment, ability to achieve the task), striving for challenging goals leads to an increase in performance of around 0.6 (±0.2) of a standard deviation on average (Locke & Latham, 1990). Research further suggests that at least four mediating processes underpin the efficacy of goal setting. From the accumulated evidence it appears that difficult and specific goals motivate people to: (1) exert more effort (e.g., work faster or harder) during task performance, (2) continue working on the task until the performance goal is reached (display persistence), (3) direct their attention to behavior and outcomes that are relevant for goal attainment (i.e., focus on the task at hand), and (4) develop and use appropriate task strategies (i.e., engage in planning; see Locke, 2000; Wegge, 2001).

However, the application of goal setting in daily work life raises a few as yet unanswered questions (Locke & Latham, 2006). For example, numerous studies have repeatedly demonstrated speed-accuracy trade-offs in performance. If goal setting using a single specific difficult goal (e.g., a speed goal) led only to a focusing effect, i.e., performance in this one aspect is maximized, while other aspects (e.g., quality) are neglected, caution would be necessary in the everyday use of goals. The problem of potential speed-accuracy trade-offs due to goal setting is particularly evident in security sensitive domains. Imagine that on a very busy day at an airport supervisors of Security Officers (SOs) would, subject to pressure from the operational business, set a goal for their SOs to “check at least 200 passengers per hour per control lane.” This would most certainly lead the SOs to a shift in behavior away from scrutinizing every passenger and bag with high caution towards faster work routines. It is highly questionable if something like this would be justifiable as security might suffer.

In airport security control, several objectives, out of which possible goals can be extracted, exist side by side. As Wetter, Lipphardt, and Hofer (2010, p. 301) stated, “The ideal security control would be effective, cheap, and would not interfere with the operation of the airport.” In other words, SOs should provide security by scrutinizing passengers and baggage, do so at a reasonable pace, not creating costly delays of flights, and, last but not least, be friendly and helpful to passengers. Thus, the coexistence of several goals, which are often not addressed and prioritized explicitly by management, can lead to goal conflicts for employees (Schmidt & Dolis, 2009; Schmidt, Kleinbeck, & Brockmann, 1984). The question arises how goal setting affects these different performance aspects and how unwanted trade-offs could be avoided. Our research aims to analyze this problem in detail, comparing the effectiveness of dual goals for speed and accuracy to single speed goals on the same task for the first time. Moreover, we also seek to identify the underlying stages of information processing (e.g., scanning in working memory, decision processes) that are affected by single and dual goals in order to gain new insights about the processes that explain the impact of goals on speed-accuracy trade-offs.

3.1.A Goal Setting and Human Information Processing

In view of the very robust performance effects linked to goal setting it can be expected that high work motivation induced by setting difficult goals should accelerate and modulate many basic processes that are necessarily involved in performing tasks. These basic processes might include, for example, the *identification* and *semantic classification* of new information or the *processing* of information in working memory (Wegge, 2001). Moreover, it can be assumed that influencing these processes is achieved through changes of activation or inhibition processes in the neural system. In support of this idea, prior studies clearly documented goal setting effects within tasks that allow for the identification of these basic processes by a thorough manipulation and analysis of reaction times (Wegge, 1998). For example, it was found that goal setting enhances the speed of scanning processes in working memory (e.g., examined using the Sternberg task), increases general working memory performance (e.g., examined using reading span tests), and also improves the identification and semantic classification of simultaneously presented pairs of letters (Wegge & Dibbelt, 2000). In addition, the performance enhancing effects of speed goals in studies of this kind could not be explained by a simple speed-accuracy trade-off. Even though we know of no other experiment in which the effectiveness of dual goals on the *same* task was investigated, based on this firm knowledge we predict:

Hypothesis 1: Setting difficult speed goals and setting difficult dual goals for speed and accuracy lead to an increase in speed.

3.1.B Goal Conflicts

Striving for multiple goals at the same time is pervasive in the work place, as work tasks become more complex over time (Humphrey, Nahrgang, & Morgeson, 2007) and organizations also tend to reduce costs through downsizing, so that the responsibilities of remaining employees are increased. Therefore, research on goal setting has also addressed the question of how goals should be formulated in situations where employees have to perform more than one task (e.g., Pritchard, Harrell, DiazGranados, & Guzman, 2008; Schmidt & Dolis, 2009; Schmidt et al., 1984). Based on this research, it can be concluded that striving for multiple or even conflicting goals is problematic and often yields performance decrements, in particular on tasks or performance aspects that the goal does not explicitly address. One possible explanation for the difficulty of realizing all aims at the same time (and the stress created by doing so) is the limited overall capacity of cognitive components, such as working memory (Hasher & Zacks, 1988) or volition (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Hagger, Wood, Stiff, & Chatzisarantis, 2010). Due to this limited capacity and given the premise of (nearly) full utilization of cognitive resources during a difficult task, focusing on one aspect of performance that is stressed in the specific performance goal (e.g., speed) must be associated with a decrease of performance in another aspect that lies outside the focus (e.g., accuracy). We thus predict:

Hypothesis 2: Setting a single, specific, difficult speed goal triggers a speed-accuracy trade-off.

If this assumption is true, the question arises as to how this potentially unwanted trade-off could be avoided. One possibility could be setting a dual goal that addresses both speed and accuracy at the same time. Performance on controlled memory search tasks like the Sternberg task (Sternberg, 1966, 1969, 1975) is strongly dependent on load (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). With a dual goal, two goals compete for the same pool of cognitive resources. Experiments by Wegge (1998) using the Sternberg paradigm showed that the serial comparison stage (representing scanning processes) is particularly affected by setting speed goals. Based on this knowledge, we predict:

Hypothesis 3: Setting a dual goal slows down scanning processes in difficult trials.

We will test our hypotheses using two different paradigms: Experiment I applies the Sternberg paradigm, and Experiment II applies the d2 test (Brickenkamp, 1962).

3.2 Experiment I

3.2.A Method

For Experiment I, the Sternberg task was chosen because it follows the same principles as the baggage screening task performed by SOs at an airport or a sports stadium. It requires participants to decide if a target stimulus was or was not contained in a stimulus set that had been learned previously. Similarly, SOs learn what items are prohibited on an airplane or in the sports stadium. Later, during their daily work, they search bags either by hand or by judging scans and decide if one of those learned stimuli is present or not. Another example in security is access control, where a person seeking entry has to be matched with a previously learned set of persons that are granted access. In another domain, namely quality control at an assembly belt, errors (deviations from a previously learned pattern) have to be identified. All these jobs have furthermore in common that processing time must not be neglected.

The Sternberg paradigm allows for analysis of different stages of mental processing by systematic variation of certain experimental factors (see below). In this experiment, participants were presented a set of one-digit numerals (e.g., 4 5 7 9) to be memorized as the first stimulus. These numerals were selected by a computer. Display duration (in s) was equal to the number of numerals contained. After a certain time interval (ca. 1 s) participants were presented a single numeral as the second stimulus (e.g., 5). Subsequently, they had to decide if the second stimulus was a part of the learned set or not. Participants indicated their response (“yes” or “no”) by pressing one of two adjacent keys on a keyboard. Participants were then presented the next set to be learned (varied-set procedure).

Setting and Participants. The setting was devised to maximize standardization; it always featured the same room with the same equipment. At the beginning of the experiment, the experimenter chose which one of the two keys on the keyboard represented the answer “yes.” Participants were instructed to use their left index finger for pressing the left key and their right index finger for pressing the right key (counterbalanced across conditions). Stimuli were presented on a CRT display. The total sample consisted of 83 students at a large European university. After the matching of the participants (see below), the data on 57 participants (70.18% men), ages 18-37 years (*M* = 25.21, *SD* = 3.91), remained for analysis. All participants had normal or corrected-to-normal vision.

Experimental Design. We manipulated four independent variables, three of which being within-subject factors and one being a between-subjects factor. The first within-subjects factor “Block” had two levels, “baseline” and “goal setting.” In the baseline block, no goals were assigned to participants. In the goal setting block, difficult, specific goals were assigned to all participants except those assigned to the control condition. The second within-subjects factor was “Set size,” the number of one-digit numerals to be learned. This factor had five levels, ranging from set size 2 to set size 6. The third within-subjects factor “Response type” had the two levels “yes” and “no.” This factor indicated the response given by the participants. Both the second and the third factor were included in Sternberg’s (1969) stage model of item recognition. According to him, manipulating set size affects serial comparison processes (longer RTs for larger set size due to more serial comparisons), whereas manipulating response type affects the binary decision process (longer RTs for “no” answers due to different decisional process).

The between-subjects factor “Goal” divided the participants into three groups. The first group served as controls. Participants were asked to “do their best” in solving the task. In combination with the two blocks, this design allowed us to assess and control for learning effects (increase in performance between baseline and goal setting block for controls). The second group received a speed goal as the instruction in the goal setting block; they were told that the mean RT in the goal setting block should be 12% lower than in the baseline block. The third group received a speed goal *and* an accuracy goal as the instruction in the goal setting block. This goal stated that the mean RT in the goal setting block should be 12% lower than in the baseline block and that no more than two additional errors (compared to the baseline) should be made.

Since numerous studies suggested that goal commitment is a moderator for goals to be effective (e.g., Erez & Kanfer, 1983; Latham & Yukl, 1975; Locke, 1968; Locke & Latham, 1984), we decided to boost goal commitment in the experiment.

According to Locke, Latham, and Erez (1988), external rewards are important determinants of commitment. Making use of this knowledge, we informed the participants that they would be awarded an equivalent of USD 10.25 for goal attainment.

Dependent variables. The computer recorded reaction times (RTs) and error rates. RT is the time span from presentation of the second stimulus (the probe) up to the participant's keystroke. In order to measure the success of the goal commitment manipulation, all participants (except controls) filled out a questionnaire comprising 9 items, using 4-point answering scales ranging from low (1) to high (4). These items assessed different antecedents of goal commitment that previous studies identified as important (see Klein, Wesson, Hollenbeck, & Alge, 1999): the value of goal attainment (e.g., "reaching this goal will have pleasant consequences for me"), expectancy of goal attainment (e.g., "It is highly likely that I am a little bit better than the goal requires"), and volitional strength during goal attainment ("I will stick to the goal, even if I find that my feelings are diverting me from this goal"). This commitment scale was found to be highly reliable (9 items; Cronbach's $\alpha = .81$).

Procedure. At the beginning of the experiment participants received standardized written instructions explaining the Sternberg task and the course of the experiment. Participants started by completing a warm-up block (120 trials with feedback after every 30 trials) that aimed at familiarizing them with the task. The main part of the experiment started with the baseline block, which consisted of 240 trials. "Set size" and "Response type" were varied systematically, and presentation order was random. After every 30 trials a feedback message showed the mean RT of the last 30 trials (correct answers only) and the number of errors. In addition, after the baseline block a feedback message showed the overall mean RT and the overall number of errors of all 240 trials.

After the participants completed the baseline block, they were matched and assigned to a goal condition by the computer using a predefined algorithm. The first 10 participants were assigned as controls (“do your best” goal). For all subsequent participants, the following matching process was followed: It was checked if the participant’s mean RT of the baseline block was comparable (i.e., ±20 ms) to the RT of one of the controls. If this was the case, they formed a matched pair, and the second participant was assigned the speed goal. If the RT was not comparable (i.e., the difference exceeded 20 ms), this participant was assigned as a control. If, in the course of the experiment, a third participant’s RT of the baseline block was comparable to the RTs of an already matched pair (i.e., ±20 ms with both other participants), this third participant was assigned the dual speed and accuracy goal and completed the trio of participants. With the help of this yoked design, baseline performance should not differ across experimental conditions. Before the start of the next block, participants (except controls) completed the questionnaire on goal commitment. The subsequent goal setting block was identical to the baseline block with one exception: feedback was different. In the goal setting block, only the overall feedback message at the end of the block was displayed, showing the overall mean RT for correct answers and the number of errors of all 240 trials. In order to compensate for the now missing intermediate feedback, a short break was inserted after every 60 trials.

3.2.B Results for Experiment I

Control Variables: Sex, Age, Commitment. Table 1 (see next page) shows the intercorrelations between variables. Male participants were older than female participants. There was a positive correlation between age and speed in both the baseline and the goal setting blocks, with younger participants being faster. As there were no significant differences in the distribution of the participants’ age between the three goal conditions (no goal: M = 26.37, SD = 3.29; speed goal: M = 25.74, SD =

4.13; speed and accuracy goal: M = 23.53, SD = 3.88) with χ^2 (2, N = 57) = 4.57, p = .10, this variable was subsequently regarded as evenly distributed between groups. The goal commitment scale revealed no difference between the group with the speed goal (M = 3.0, SD = 0.5) and the group with the dual speed and accuracy goal (M = 2.8, SD = 0.5) with $t(36)$ = 1.20, p = .24. These mean values further suggested that overall goal commitment was given and constant across goal conditions.

Table 1

Experiment I: Pearson correlations for age, sex, and performance measures.

				Baseline		Goal setting
	Age	Sex	Commitment	Speed	Accuracy	Speed
Sex[2]	**-.48*** **					
Commitment	.20	-.27				
Speed in baseline block	**.35** **	-.25	.15			
Accuracy in baseline block	-.12	.16	-.10	-.02		
Speed in goal setting block	**.34** **	-.25	-.05	**.82*** **	-.09	
Accuracy in goal setting block	-.17	.01	-.07	-.15	**.53*** **	**-.30***

Note. Significant correlations are shown in **bold** (* p < .05; ** p < .01; *** p < .001).

Model Fit. To analyze whether the Sternberg model can be applied to the present data, a mixed MANOVA with the between-subjects factor “Goal” (do your best, speed, speed and accuracy) and the within-subjects factors “Block” (baseline, goal setting), “Response type” (yes, no), and “Set size” (2, 3, 4, 5, 6) on the dependent variable RT (ms) was calculated (Tables 2 and 3, see next pages). This MANOVA

[2] Male is coded as 0, whereas female is coded as 1.

revealed significant main effects of “Block,” “Response type,” and “Set size,” as predicted by the Sternberg model. Significant interactions were “Goal x Block,” “Block x Set size,” “Response type x Set size,” and “Block x Response type x Set size.”

Table 2

Performance measures Experiment I: means (standard deviations) for RTs in ms, above, and errors in %, below.

	Set size 2			Set size 3			Set size 4			Set size 5			Set size 6		
Resp. type	Baseline	Goal setting	Δ (GS-BL)	Baseline	Goal setting	Δ (GS-BL)	Baseline	Goal setting	Δ (GS-BL)	Baseline	Goal setting	Δ (GS-BL)	Baseline	Goal setting	Δ (GS-BL)
Goal 0: no goal															
yes	506 (74) 4.2 (4.8)	467 (74) 5.3 (3.9)	-39 (47) 1.1 (4.1)	563 (84) 2.4 (4.0)	512 (86) 1.1 (2.3)	-51 (47) -1.3 (5.0)	630 (100) 2.6 (2.9)	560 (95) 3.3 (3.6)	-70 (46) 0.7 (3.7)	665 (119) 4.2 (3.9)	611 (118) 5.7 (5.2)	-54 (45) 1.5 (5.0)	672 (122) 3.1 (3.1)	628 (131) 5.0 (5.1)	-44 (60) 2.0 (5.3)
no	557 (82) 2.0 (2.6)	524 (69) 2.2 (2.9)	-33 (54) 0.2 (3.5)	597 (91) 1.1 (2.3)	579 (107) 1.8 (3.2)	-18 (52) 0.7 (4.4)	646 (110) 1.3 (2.0)	603 (105) 1.8 (2.1)	-44 (48) 0.4 (2.8)	701 (139) 2.0 (3.2)	679 (155) 1.8 (4.0)	-22 (60) -0.2 (2.6)	743 (163) 5.1 (1.7)	656 (131) 5.3 (3.0)	-86 (79) 0.2 (2.6)
Goal 1: speed goal															
yes	506 (74) 5.3 (4.6)	439 (69) 7.7 (6.8)	-67 (45) 2.4 (6.6)	568 (83) 2.4 (3.5)	468 (64) 4.8 (5.9)	-100 (53) 2.4 (6.2)	631 (97) 3.3 (3.3)	505 (70) 5.9 (5.3)	-127 (59) 2.6 (5.8)	660 (98) 3.7 (4.1)	540 (71) 6.1 (7.5)	-120 (63) 2.4 (6.6)	678 (112) 5.3 (5.5)	572 (97) 7.9 (7.7)	-107 (76) 2.6 (6.1)
no	556 (88) 1.1 (2.3)	471 (71) 5.5 (7.4)	-85 (51) 4.4 (7.4)	591 (87) 0.9 (2.2)	505 (81) 2.9 (4.2)	-85 (52) 2.0 (3.5)	634 (121) 1.5 (3.2)	521 (87) 1.5 (3.2)	-112 (73) 0.0 (3.4)	697 (133) 1.3 (2.4)	595 (99) 3.3 (3.3)	-101 (73) 2.0 (3.5)	712 (147) 6.2 (3.5)	608 (141) 9.4 (4.8)	-104 (80) 3.3 (5.8)
Goal 2: speed & accuracy goal															
yes	523 (111) 4.2 (4.2)	450 (53) 5.5 (6.7)	-73 (89) 1.3 (5.4)	569 (78) 1.3 (2.4)	486 (54) 3.7 (6.2)	-83 (38) 2.4 (4.9)	641 (110) 2.6 (4.0)	533 (69) 4.0 (6.7)	-108 (91) 1.3 (8.2)	658 (105) 3.7 (3.4)	573 (72) 5.3 (5.3)	-85 (66) 1.5 (6.7)	679 (120) 4.2 (4.4)	606 (104) 6.1 (5.3)	-74 (70) 2.0 (5.1)
no	559 (98) 1.5 (2.8)	463 (51) 4.0 (4.9)	-96 (72) 2.4 (4.7)	605 (96) 0.9 (1.8)	522 (64) 2.8 (5.0)	-82 (65) 2.0 (4.9)	644 (110) 1.8 (2.9)	551 (73) 2.6 (2.9)	-93 (76) 0.9 (3.6)	690 (129) 1.5 (2.5)	605 (88) 3.1 (5.2)	-85 (75) 1.5 (6.4)	701 (110) 7.9 (3.6)	613 (95) 9.2 (11.2)	-88 (77) 1.3 (11.7)

Table 3

Performance measures Experiment I: Test statistics, significance probabilities, and effect sizes.

	RTs	Errors
Goal	$F(2, 54) = 0.70$, $p = .50$	$F(2, 54) = 1.34$, $p = .27$
Block	**$F(1, 54) = 140.39$, $p < .001$, $\eta^2_p = .72$**	**$F(1, 54) = 14.00$, $p < .001$, $\eta^2_p = .21$**
Response type	**$F(1, 54) = 27.72$, $p < .001$, $\eta^2_p = .34$**	**$F(1, 54) = 21.99$, $p < .001$, $\eta^2_p = .29$**
Set size	**$F(4, 216) = 198.78$, $p < .001$, $\eta^2_p = .79$**	**$F(4, 216) = 45.37$, $p < .001$, $\eta^2_p = .46$**
Goal x Block	**$F(2, 54) = 6.27$, $p < .01$, $\eta^2_p = .19$**	$F(2, 54) = 1.79$, $p = .18$
Goal x Response type	$F(2, 54) = 1.10$, $p = .34$	$F(2, 54) = 2.25$, $p = .12$
Goal x Set size	$F(8, 216) = 0.38$, $p = .93$	$F(8, 216) = 1.68$, $p = .11$
Block x Response type	$F(1, 54) = 1.32$, $p = .26$	$F(1, 54) = 0.53$, $p = .47$
Block x Set size	**$F(4, 216) = 6.06$, $p < .001$, $\eta^2_p = .10$**	$F(4, 216) = 0.93$, $p = .45$
Response type x Set size	**$F(4, 216) = 4.58$, $p < .01$, $\eta^2_p = .08$**	**$F(4, 216) = 11.41$, $p < .001$, $\eta^2_p = .17$**
Goal x Block x Response type	$F(2, 54) = 1.38$, $p = .26$	$F(2, 54) = 0.14$, $p = .87$
Goal x Block x Set size	$F(8, 216) = 1.79$, $p = .08$	$F(8, 216) = 0.55$, $p = .82$
Goal x Response type x Set size	$F(8, 216) = 0.55$, $p = .82$	$F(8, 216) = 0.41$, $p = .91$
Block x Response type x Set size	**$F(4, 216) = 6.12$, $p < .001$, $\eta^2_p = .10$**	$F(4, 216) = 0.63$, $p = .64$
Goal x Block x Response type x Set size	$F(8, 216) = 1.59$, $p = .13$	$F(8, 216) = 0.69$, $p = .70$

In a similar way, the conformity of the data with the Sternberg model was tested for accuracy. A mixed MANOVA with the between-subjects factor “Goal” (do your best, speed, speed and accuracy) and the within-subjects factors “Block” (baseline, goal setting), “Response type” (yes, no), and “Set size” (2, 3, 4, 5, 6) on the dependent variable errors (percent) was calculated (Tables 2, 3). This MANOVA revealed significant main effects of “Block,” “Response type,” and “Set size,” as predicted by the Sternberg model. There was also a significant interaction “Response type x Set size.”

A significant interaction “Response type x Set size” means that these factors at least partly influence the same stage of information processing. This is contrary to what Sternberg (1969) had postulated and found – namely, that the factors are completely additive, i.e., independent of each other. This interaction is solely based on RT differences between “yes” and “no” answers being larger in easy tasks (set size 2) than in difficult tasks (set size > 2^3). Even though this implies that in our experiment some caution is advised when interpreting the effects of the binary decision stage ("Response type") and the serial comparison stage ("Set size") on their own, the expected independence for other levels of these factors is given.

Hypothesis Testing. The significant interaction "Goal x Block" in the MANOVA on RTs (Tables 2, 3) suggests that goal setting was effective. A closer look at the data (Figure 1) and deeper analyses with simple contrasts revealed significant differences in the mean RT for no goal vs. speed (t = -3.41, p < .01), and for no goal vs. speed and accuracy (t = -2.53, p < .05) but no significant difference for speed vs. speed and accuracy (t = -0.88, p = .38). Thus, both goals were effective in reducing RTs and Hypothesis 1 is supported by the data.

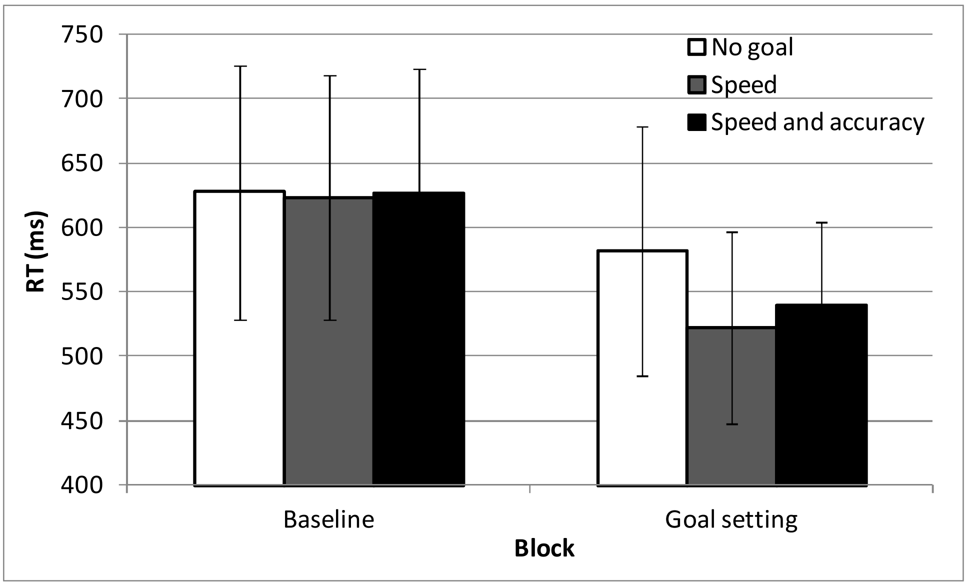

Figure 1. Effect of "Goal" x "Block" on RTs (means and standard deviations) in Experiment I.

Table 1 shows that in the baseline block (N = 57), speed and accuracy were uncorrelated (r = -.02, p = .87), whereas in the goal setting block (N = 57), they were correlated (r = -.30, p < .05). Subsequently, the data of the goal setting block were

[3] I.e., the average of set sizes 3 to 6.

analyzed separately for the different goal conditions (each with $n = 19$). The group with the “do your best” goal reached a correlation of $r = -.29$, the group with the speed goal reached a correlation of $r = -.41$, and the group with the dual goal reached a correlation of $r = -.14$. The correlation is highest for the group with the speed goal but does not reach significance ($p = .08$) due to the relatively low power. It is thus evident that goal setting triggered an *overall* speed-accuracy trade-off, to which the speed goal contributed more than the dual goal did. Hypothesis 2 therefore also finds support in the data.

As the above reported results of the MANOVAs are tests averaged across all levels of the factors, an in-depth analysis on the factor “Set size” representing scanning processes was performed. Simple contrasts for the factor “Goal” and Helmert contrasts for the factor “Set size” on RT (ms) were calculated. For the interaction “Goal” x “Block” x “Set size” (2 vs. >2), the contrasts indicated a significant difference for speed vs. speed and accuracy, with $t = 3.02$, $p < .01$ (Table 4, see next page). For the simple speed goal, improvements due to goal setting were much higher in difficult sub-tasks (set size > 2) than in simple sub-tasks. For the dual goal these differences were no longer observable, because scanning is slowed down again in difficult sub-tasks. It can be concluded that the dual goal particularly slowed down scanning processes in the more difficult trials, thus supporting Hypothesis 3.

Table 4

Experiment I: Interaction Goal x Block x Set size. Means (standard deviations) for RTs in ms, above, and errors in %, below.

	Goal 1: speed goal			Goal 2: speed & accuracy goal		
	Baseline	Goal setting	Δ (GS-BL)	Baseline	Goal setting	Δ (GS-BL)
Set size 2	531 (78)	455 (66)	-76 (44)	541 (100)	456 (47)	-85 (76)
	3.2 (2.5)	6.6 (5.9)	3.4 (5.6)	2.9 (2.4)	4.7 (4.9)	1.9 (4.4)
Set size 3	579 (80)	487 (68)	-93 (46)	587 (84)	504 (55)	-83 (47)
	1.7 (2.2)	3.8 (4.2)	2.2 (3.8)	1.1 (1.6)	3.3 (4.7)	2.2 (4.1)
Set size 4	632 (103)	513 (74)	-119 (59)	642 (107)	542 (68)	-101 (78)
	2.4 (2.2)	3.7 (3.6)	1.3 (3.3)	2.2 (2.8)	3.3 (4.3)	1.1 (5.1)
Set size 5	678 (111)	568 (77)	-110 (62)	674 (111)	589 (77)	-85 (64)
	2.5 (2.4)	4.7 (4.3)	2.2 (3.7)	2.6 (2.1)	4.2 (4.8)	1.5 (5.4)
Set size 6	696 (123)	589 (113)	-106 (71)	690 (112)	609 (96)	-80 (70)
	5.7 (3.7)	8.7 (5.3)	2.9 (4.3)	6.0 (2.8)	7.7 (6.8)	1.7 (6.3)

Note. Helmert contrasts on RTs (significant results are shown in **bold**): **Set size 2 vs. >2: t = 3.02, p < .01;** Set size 3 vs. >3: t = 0.99, p = .33; Set size 4 vs. >4: t = 0.53, p = .60; Set size 5 vs. 6: t = -0.06, p = .95.

3.2.C Discussion of Experiment I

By applying the Sternberg paradigm, it could be shown that both single and dual goals were effective in increasing speed performance. The single goal was not significantly more effective than the dual goal. However, the single specific difficult goal triggered a speed-accuracy trade-off, which was less pronounced with the dual goal. Our results further showed that the joint optimization of speed performance and accuracy performance by dual goals was achieved by slowing scanning processes in difficult trials with high cognitive load. The evidence created by Experiment I was gained in a highly controlled setting with one task and a rather homogeneous sample of students as participants. In order to investigate if these findings generalize to another task and a different sample of participants, Experiment II was devised.

3.3 Experiment II

3.3.A Method

For Experiment II, the d2 test was chosen. Whereas the Sternberg task is a number memorization task, the d2 is a test of attention consisting of a visual discrimination task (targets have to be discriminated from distracters, e.g., ḋ from d̈). The d2 task also shares some basic principles with tasks in security. For example, a visual ticket check regarding validity date at a checkpoint could be seen as a similar task: Holders of valid tickets (targets) have to be efficiently granted access, which has to be reliably denied to holders of invalid tickets. Given the usual queues at checkpoints and people arriving rather late than early for boarding of the aircraft or the sports game, there is considerable time pressure on these checks, which is reflected in the concept of the test as well.

Whereas Experiment I featured a typical student sample, Experiment II featured a more heterogeneous sample of participants which is likely to be much more representative of the general working population. 19 participants (57.89% female), ages 23-61 years (M = 38.16, SD = 15.18) with different educational backgrounds and different jobs volunteered. All participants had normal or corrected-to-normal vision. The d2 was administered as paper and pencil test.

Participants were tested three times in a repeated measures design. In order to lessen learning and memory effects, a time gap of one month was inserted after each test. Thus, the three tests were completed in three months time, with one test per month. The first testing served as baseline; participants were told to “do their best.” The d2’s speed (total number of symbols processed) and accuracy (total number of errors) measures were recorded as dependent variables. For the second testing, participants were matched to two groups that did not differ significantly with regard to their baseline performance in speed and accuracy (all *p*s > .82). Before the second testing, participants of group 1 were told that they should completely solve every line of the test in the given time, whereas participants of group 2 were told that they should not commit any errors. This goal manipulation represents a specific difficult goal regarding speed (group 1) or accuracy (group 2). The third testing used a dual goal; all participants were told that they should completely solve every line of the test in the given time *and* not commit any errors.

The procedure was as follows: After having arrived, participants were explained the d2 task. They were then assigned the goal and, subsequently, solved the test. After having solved the test, they were asked by the experimenter to repeat the goal they had been assigned. This question served as manipulation check. One participant mentioned that the goal had not at all been present in her mind throughout the testing. Her results were subsequently excluded. 11 participants (42%) did not complete all three tests. The reasons for dropping out before the last testing were lack of motivation (9), severe illness (1), and absence (1).

3.3.B Results for Experiment II

Performance measures are presented in Table 5 (see next page) and visualized in Figure 2 (see next but one page). Results from group 1 reveal that speed performance (assessed by the number of symbols processed by participants) with the specific difficult speed goal was higher than do your best with $t(6) = 3.00$, $p < .05$. Similarly, speed performance with the dual goal was higher than do your best performance with $t(4) = 3.84$, $p < .05$. Speed performance with the dual goal appeared to be slightly lower than with the speed goal, although this result did not reach significance (p = .08). Thus, both goals were effective in increasing speed of performance and Hypothesis 1 is supported again.

Table 5

Performance Measures Experiment II: means (standard deviations) for symbols processed, top, errors, middle, and speed-accuracy correlation.

	Group 1 (specific difficult = speed)	Group 2 (specific difficult = accuracy)
Do your best	479 (119)	490 (93)
	38.1 (42.0)	42.6 (43.9)
	r = .76	r = .61
	n = 9	n = 10
Specific difficult goal	547 (70)	485 (101)
	27.3 (23.4)	17.8 (21.1)
	r = .42	r = .61
	n = 7	n = 9
Dual goal	534 (88)	540 (99)
	17.8 (17.9)	18.7 (23.8)
	r = .24	r = .37
	n = 5	n = 6

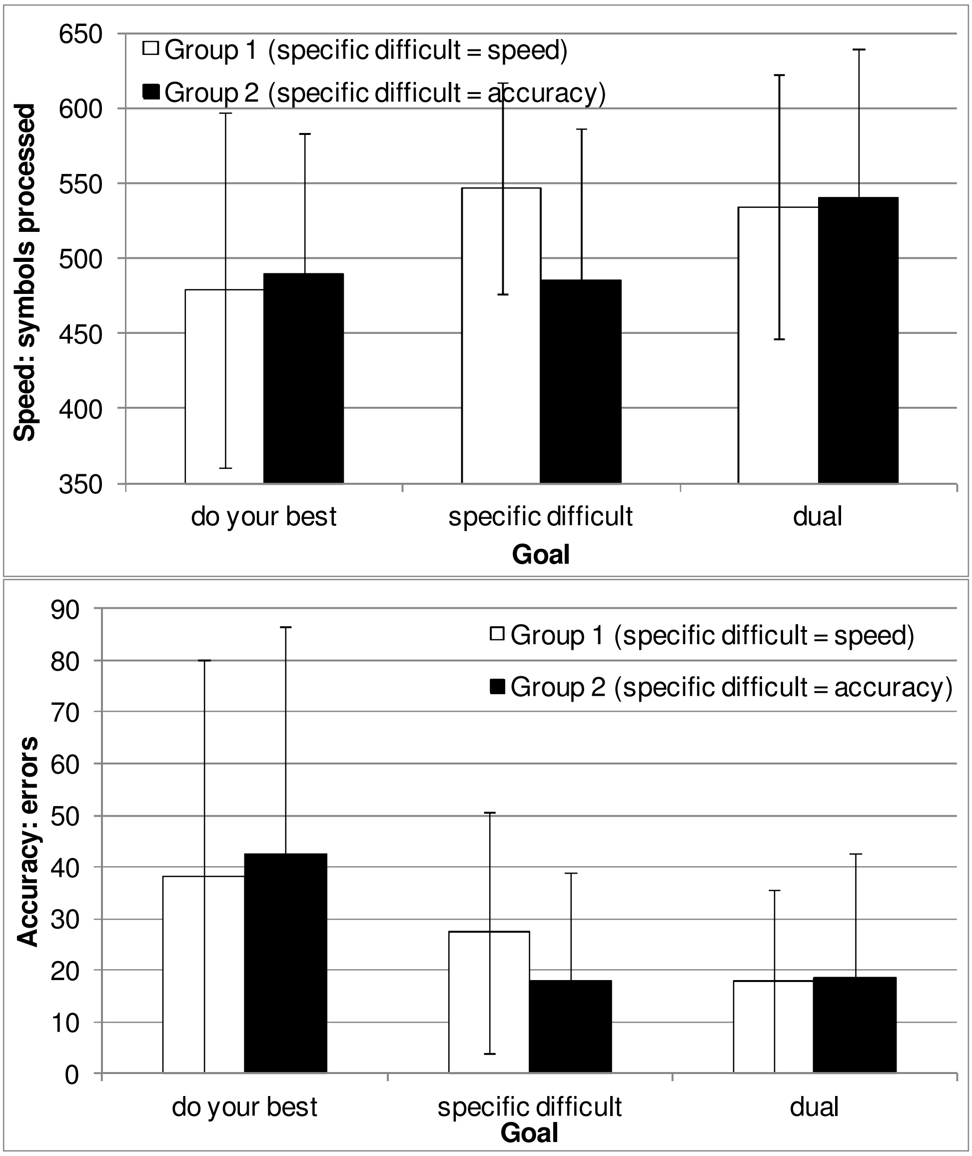

Figure 2. Performance measures Experiment II: Means (standard deviations) for speed, above, and accuracy, below.

Table 5 shows that for "do your best" and the specific difficult goal conditions, speed and accuracy were strongly correlated, indicating that participants paid with more errors for being faster (in this task, positive correlations indicate speed-accuracy trade-offs). Even though none of these correlations reaches significance due to the low power (all *p*s > .13), these data are in line with Hypothesis 2. With the dual goal, the observed speed-accuracy trade-off is again less pronounced. However, due to the low

N, this experiment cannot provide conclusive evidence on its own and thus has to be interpreted together with Experiment I.

3.4 General Discussion

By applying the Sternberg paradigm and the d2 test in two experiments with different samples, this study shows that both single and dual goals are effective in increasing working speed. However, the single specific difficult goal triggered a speed-accuracy trade-off, which was less pronounced with the dual specific difficult goal. The joint optimization of speed *and* accuracy performance by dual goals was achieved by slowing scanning processes in difficult trials. The implications of these findings for practical applications in the area of goal setting are manifold, especially in domains with multiple objectives or in security sensitive domains where security sometimes has to be provided under time pressure. We suggest that a dual goal might be used if speed-accuracy trade-offs are to be prevented. This might work best with routine tasks with little cognitive load.

Many studies over the last 50 years demonstrated that single specific difficult goals are effective (Locke & Latham, 2002). Based on our results, which revealed a speed-accuracy trade-off using this type of goal, we would like to argue that a part of this “increase in performance” described in many studies on goal setting is due to a loss of performance in another work aspect. Single goals apparently have the effect of “guiding” and “focusing” performance and effort on one aspect. According to Shiffrin and Schneider (1977), error rates might go up when time pressure is induced, because the participants might be forced to terminate the controlled search before it is completed. Then, a response would have to be made as a guess, based on partial information.

More important, we are not aware of a previous study that directly contrasted the effects of a single goal to a dual goal on task performance. By doing so, we found that the dual goal was effective in reducing RTs as well, with a slightly less pronounced side-effect of triggering a speed-accuracy trade-off. When task difficulty and the resulting cognitive load were high, scanning processes were slowed down. In their research on the effects of dual goals on self-regulation and resource allocation, Schmidt and Dolis (2009) suggested that the cumulative demands placed by multiple difficult goals may exceed individuals' perceived capabilities and may lead to partial or total abandonment of one goal to ensure attainment of the other. Linking these ideas with our results on scanning processes, we would like to argue that there are two possible processes that can explain these results: (1) Individuals can consciously assign lower importance to one goal, which might ultimately lead them to give it up entirely, or (2) they can keep following both goals consciously, but unconscious cognitive processes (e.g., scanning processes) that are affected by goal setting and task demands limit the outcome in one goal area.

Attempting to achieve speed and attempting to achieve accuracy can be seen as two things that interfere with each other, competing for a general limited capacity (Kahneman, 1973; Moray, 1967). In Kahneman's (1973) words, "it is easy to focus attention exclusively on a single object and difficult to divide attention among several objects" (p. 7). In addition, it is known that controlled search tasks like the Sternberg task are limited-capacity processes that require attention and are thus strongly dependent on load (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). By allocating capacity to two different aspects – namely, speed and accuracy, the dual goal actually divided the available capacity. Our research thus supports the conclusion by Schmidt et al. (1984) that goals control the allocation policies for the resources invested.

Our results further hint at the importance of task characteristics (which determine task difficulty) in relation to the limited capacity for the actual task performance, if more than just one aspect is taken into account (e.g., speed and accuracy). Thus, in our view, a general assumption of "something's got to give" in a dual goal (or even dual task) situation is wrong, because the occurrence of trade-offs depends on the relation between task demands and available capacity. It can be assumed that available capacity can change (e.g., during the day). Schmidt, Kleinbeck, and Knauth (1988) have demonstrated that setting goals that are adapted to the human circadian rhythm can indeed contribute to an optimal distribution of work units and rest periods in the real work situation. In that study, this optimization of allocation policies lowered subjective strain of employees.

To integrate findings, it could be said that single goals can have the disadvantage of inducing trade-offs. The advantages of single goals are that they are well known, easy to use (also with difficult tasks), and the most effective. Dual goals, on the other hand, seem to slightly reduce trade-offs but are somewhat more complicated to use and understand and are a little less effective than single goals. At present there is not much knowledge available on dual goals. What type of goal should be used ideally depends largely on the overall job characteristics (e.g., if there is one aspect that must not be neglected at all) but also on task difficulty (e.g., routine vs. non-routine task).

It was not possible to conduct this study in the field because performance in security areas cannot be compromised. Therefore, no final conclusion can yet be drawn about the effects of dual goals in the everyday setting in security. As a next step in order to pave the way for applications in security sensitive domains, we suggest a simulator study (e.g., using X-Ray Tutor, Schwaninger, 2003, 2004). Thus, real stimulus material (e.g., x-ray images with and without prohibited items) could be used to get closer to reality and to see if the effects can be replicated. In particular, it should be tested to what extent dual goals have the potential to reduce trade-offs. If trade-offs can be minimized with dual-goals, a first pilot test in the real work situation could be ventured. On conducting this pilot test, it would be necessary to find ways of reliably measuring performance of employees in different aspects. It should further be taken great care to ensure as much experimental control as possible and to rule out as many confounds as possible. We also suggest that the dual goal should contain one component about security, so that it can be ensured that employees keep this crucial aspect in focus as well.

In most work environments, several goals exist side by side, so it is of high practical relevance to find ways of dealing optimally with this situation. Apart from defining key performance indicators and deriving goals, the goals themselves and the tasks to be performed have to be analyzed regarding their difficulty and capacity demands. If there are several goals to strive for at the same time or several tasks to work on concurrently, they should be integrated into a system. Tasks and goals should be prioritized and minimally acceptable values of key performance indicators defined. This would give employees a clearly defined system of how to work in their complex environment without being overwhelmed by different goals, demands, and tasks. Coming from a different starting point, Schmidt and Dolis (2009) reached the same conclusion – namely, that a system that promotes communication and coordination of goals may help to avoid overloading or underloading of employees. A system of this kind can be set up by applying ProMES (Pritchard, 1990), for example, which is an approach that seeks to support group members in solving these problems (Wegge et al., 2010). As Wetter, Fuhrmann, Lipphardt, and Hofer (2011) showed in a pioneering study at the security control of a large European airport, this can be a fruitful approach in security domains, where many stakeholders are involved and performance in certain aspects must not be compromised.

3.5 References

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social psychology, 74,* 1252-1265.

Brickenkamp, R. (1962). *Test d2: Aufmerksamkeits-Belastungs-Test. [Test d2: Test of attention and stress].* Göttingen: Hogrefe.

Erez, M., & Kanfer, F. H. (1983). The role of goal acceptance in goal setting and task performance. *Academy of Management Review, 8,* 454-463.

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin, 136,* 495-525.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Vol. 22* (p. 196). San Diego, CA: Academic Press, Inc.

Humphrey, S. E., Nahrgang, J. D., & Morgeson, F. P. (2007). Integrating motivational, social, and contextual work design features: A meta-analytical summary and theoretical extension of the work design literature. *Journal of Applied Psychology*, *92*, 1332-1356.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Klein, H. J., Wesson, M. J., Hollenbeck, J. R., & Alge, B. J. (1999). Goal commitment and the goal-setting process: Conceptual clarification and empirical synthesis. *Journal of Applied Psychology, 84,* 885-896.

Latham, G. P., & Yukl, G. A. (1975). A review of research on the application of goal setting in organizations. *The Academy of Management Journal, 18,* 824-845.

Locke, E. A. (1968). Towards a theory of task motivation and incentives. *Organizational Behavior and Human Performance, 3,* 157-189.

Locke, E. A. (2000). Motivation, cognition and action: An analysis of studies on task goals and knowledge. *Applied Psychology: An International Review*, *49*, 408-429.

Locke, E. A., & Latham, G. P. (1984). *Goal setting for individuals, groups, and organizations.* Chicago, IL: Science Research Associates.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance.* Englewood Cliffs, NJ: Prentice Hall.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. *American Psychologist, 57,* 705-717.

Locke, E. A., & Latham, G. P. (2006). New directions in goal setting theory. *Current Directions in Psychological Science*, *15*, 265-268.

Locke, E. A., Latham, G. P., & Erez, M. (1988). The determinants of goal commitment. *Academy of Management Review, 13,* 23-39.

Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica, 27,* 84-92.

Pritchard, R. D. (1990). *Measuring and improving organizational productivity: A practical guide.* New York, NY: Praeger.

Pritchard, R. D., Harrell, M. M., DiazGranados, D., & Guzman, M. J. (2008). The Productivity Measurement and Enhancement System: A meta-analysis. *Journal of Applied Psychology*, *93*, 540-567.

Schmidt, A. M., & Dolis, C. M. (2009). Something's got to give: The effects of dual-goal difficulty, goal progress, and expectancies on resource allocation. *Journal of Applied Psychology, 94,* 678-691.

Schmidt, K.-H., Kleinbeck, U., & Brockmann, W. (1984). Motivational control of motor performance by goal setting in a dual-task situation. *Psychological Research, 46,* 129-141.

Schmidt, K.-H., Kleinbeck, U., & Knauth, P. (1988). Über die Wirkung von Leistungsvorgaben auf das Pausenverhalten und das Beanspruchungserleben bei freier Arbeit [The effect of performance goals on the distribution of rest periods and subjective strain]. *Zeitschrift für Arbeitswissenschaft, 2,* 96-101.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1-66.

Schwaninger, A. (2003). Training of airport security screeners. *Airport, 05,* 11-13.

Schwaninger, A. (2004). Computer based training: A powerful tool to the enhancement of human factors. *Aviation Security International, FEB/2004,* 31-36.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127-190.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652-654.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57,* 421-457.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27,* 1-32.

Wegge, J. (1998). *Lernmotivation, Informationsverarbeitung, Leistung.* [Academic motivation, information processing, and performance]. Münster: Waxmann.

Wegge, J. (2001). Motivation, information processing and performance: Effects of goal setting on basic cognitive processes. In A. Efklides, J. Kuhl, & R. Sorrentino (Eds.), *Trends and prospects in motivation research* (pp. 269-296). Dordrecht: Kluwer.

Wegge, J., & Dibbelt, S. (2000). Zur Wirkung von Zielsetzungen auf die Informationsverarbeitung bei Buchstabenvergleichsaufgaben [Effects of goal-setting on information processing in letter-matching tasks]. *Zeitschrift für Experimentelle Psychologie, 47,* 89-114.

Wegge, J., Jeppesen, H. J., Weber, W. G., Pearce, C. L., Silva, S. A., Pundt, A., et al. (2010). Promoting work motivation in organizations: Should employee involvement in organizational leadership become a new tool in the organizational psychologist's kit? *Journal of Personnel Psychology, 9,* 154-171.

Wetter, O.E., Fuhrmann, H., Lipphardt, M., & Hofer, F. (2011). Bringing adversaries together: The importance of a common management-level approach in complex work domains. In M. S. Faundez-Zanuy, V. Espinosa-Duró, & L. D. Sanson (Eds.), *Proceedings of the 45th International Carnahan Conference on Security Technology* (pp. 20-25). Mataró: Institute of Electrical and Electronics Engineers.

Wetter, O. E., Lipphardt, M., & Hofer, F. (2010). External and internal influences on the security control process at airports. In D. A. Pritchard, & L. D. Sanson (Eds.), *Proceedings of the 44th International Carnahan Conference on Security Technology* (pp. 301-309). San José, CA: Institute of Electrical and Electronics Engineers.

4 Bringing adversaries together: The importance of a common management-level approach in complex work domains[1]

Olive E. Wetter, Kantonspolizei Zürich, Research & Development, Zurich-Airport

Hartwig Fuhrmann, t-velopment, Dortmund

Michael Lipphardt, Flughafen Zürich AG, Terminal Engineering, Zurich-Airport

Franziska Hofer, Kantonspolizei Zürich, Research & Development, Zurich-Airport

Abstract – Ensuring smooth operation of an airport is a complex task with many stakeholders involved that have to work together closely in order to achieve optimal results. In particular, the field of security control harbors diverging interests of different entities that share responsibility for secure and timely flights. These entities usually comprise governmental organizations (e.g., legislative authorities, police) and private companies (e.g., airport operators, private security companies). Due to conflicting interests between these entities, managers might perceive each other as adversaries rather than as partners for collaboration.

To our knowledge, the present report is the first one to relate procedure and results of a fruitful collaboration on management level between representatives of these different entities at an airport, including a governmental organization as the executive, i.e., carrying out the security control, and a private company (the airport operator). In the setting of the security control at a large European airport, the Productivity Measurement and Enhancement System (ProMES) methodology [1] was

[1] This study was published 2011 in M. Faundez-Zanuy, V. Espinosa-Duró, & L. D. Sanson (Eds.), *Proceedings of the 45th Carnahan Conference on Security Technology* (pp. 20-25). Mataró, Spain: Institute of Electrical and Electronics Engineers.

chosen to identify stakeholders in the security control process as a first step in order to be able to define their interests, demands, and needs in a second step. Subsequently, justified demands were grouped and categorized into six goal domains. For each goal domain, key performance indicators were developed. Last but not least, the indicators were prioritized by weighing their relevance against each other in order to resolve conflicting interests. This process can be rather quick and efficient provided it is well prepared and monitored. It is further suggested to have an independent facilitator supervise and moderate the process.

This study highlights the importance of bringing together high level managers of different organizational entities with a systematical approach in order to pave the way for good cooperation in complex work domains. As a result, stakeholders and demands can be identified, conflicting demands can be discussed, and priorities set. Subsequently, those guidelines can be handed over to the next lower level in the organizational hierarchy. Supervisors on this next level can then define goals for their respective domains using the guidelines provided by the management.

4.1 Introduction

Offering passengers safe, secure, comfortable, and timely flights is a complex service, which demands for highly professional management. In order to offer travelers the best possible service, an airport has to rely on many interdependent stakeholders. Following the passengers' way from arrival at the airport to take-off, this becomes evident. The passengers have to arrive on time at check-in, which requires public and private transport to the airport to be fluid. Waiting-times at check-in should be within the expected range so that the necessary information and baggage can be processed in time. Ground handling companies have to ensure that baggage is loaded onto the airplane after it was checked by the police or a private security company. The passengers themselves have to find their way to the gate as conveniently as possible.

One point of particular importance is the security control, which can become a bottleneck in this chain, as all bags and passengers have to pass it. The ideal security control would be effective, cheap, and would not interfere with the operation of the airport [2]. Security control is a process that can be regarded as an excellent example showing the existing interdependencies at an airport. Security control itself is dependent on passengers. If they inform themselves prior to the flight about current regulations, arrive on time, and if they cooperate well with Security Officers (SOs), the process will most certainly be quick and without any difficulty. Security control also depends on the airport as a provider of facilities and services. One example for that is that passenger throughput differs between different docks of the same airport even when using the same security equipment [2]. Furthermore, security control also depends on the regulatory framework that can differ from country to country.

On the other side passengers, the airport, and the airlines are highly dependent on security and services provided by the security control. In a recent survey with 1315 travelers at 12 major U.S. airports [3], 79% of the passengers said that privacy is important or very important to them. 61% stated that security was even more important to them than privacy. As a third factor, efficiency seems to be important as well: 65% said they would choose a screening procedure other than scans or pat-downs if it caused a delay of only 1min vs. only 21% who said they would accept 5min delay and 9% who would accept 10min delay. Moreover, if security control does not operate smoothly, no airplane will leave the airport on time. Finally, the airport would like the security control to spread the appearance of the airport as a nice and customer-friendly place in which travelers are welcome and looked after.

These interdependencies and different interests can create tensions or an atmosphere of being adversaries rather than collaborators, although at the end of the day, the final goal of all stakeholders is the same, namely, secure, safe, comfortable and timely flights. In order to either prevent tensions from occurring or to reduce them in case they have already occurred, it is necessary to set up a systematic way of collaboration, offering consideration to all stakeholders. All parties involved can thus realize that it is necessary to create solutions together and help each other implement them. A first fruitful approach that yielded deeper insights into the security control process, carried out in close collaboration of an airport and a security control operator, was reported last year [2].

In order to foster a common management-level approach to security control, the ProMES methodology was chosen. The purpose of this paper is to document this collaboration on management-level between security control by an airport and a security control operator. Knowledge gained shall be shared, and ProMES be presented as a method for successfully setting up a productivity and performance measurement and management system in security domains.

4.2 Method

4.2.A Setting and stakeholders

The present study was carried out at a large European airport that serves as an airline hub. Two main stakeholders were involved in the project, namely, the airport operator, which is a privatized company of which the state still holds a share, and the airport police, a division of the state police. The branches involved from the airport operator were Airport Operation, responsible for smooth operation of flights, Airport Security, responsible for the airport's compliance with security standards, and Planning and Engineering, responsible for infrastructure. From the police side, the security control branch was involved, which is responsible for carrying out security control. All stakeholders assigned high-level management representatives for setting the ground for the ProMES system.

4.2.B Four ProMES workshops

In four workshops, each lasting about 4 hrs, plus 30 min of individual computer-based data collection, the management representatives set up a ProMES system referring to management level. This process was guided and supervised by an independent facilitator (the second author), assisted by the police's research & development unit (the first and the last author). Setting up a ProMES system basically consists of identifying objectives, identifying valid indicators that measure how well objectives are achieved, defining contingency functions that link indicator values to effectiveness, and setting up a system of regular measurement and feedback [1].

The main reasons for choosing ProMES were that the parties involved wanted to use a standardized and evaluated approach for identifying the different stakeholders' needs and interests, create key performance indicators, weigh their effectiveness against each other in order to resolve conflicts of interest, set up a framework of measurement, and implement regular measurement and feedback processes. The main goal of the whole project was to enhance performance and productivity, as well as employee participation and satisfaction. Starting this process on management level aimed at creating clear guidelines for the next lower levels in the hierarchy, so that these supervisors would then clearly know which goal domains the management level considers as highly important. These insights could later on be implemented in subsequent ProMES systems on these lower organizational levels.

4.3 Procedure

4.3.A Workshop I: Customers, objectives, goal domains

In the first workshop, different groups of customers for the security control were identified, and their demands and expectations listed. Subsequently, a consensus was reached about which of those demands were regarded as justified. Next, other objectives and activities reflecting the airport's key interests were added. Those objectives and justified demands were matched with groups of customers and duplications removed. The resulting objectives (n = 21) were grouped and arranged into meaningful goal domains with regard to content. These goal domains were illustrated using a mind map [4]. The results were then checked again to ensure completeness. In order to do so, the four perspectives from the Balanced Scorecard were applied: financial, customer, internal-business-process, as well as learning and growth perspective [5, 6]. The participating managers were asked to check if every perspective was contained in the collection of objectives and that none was lacking. For each of the four perspectives, objectives were contained.

4.3.B Workshop II: Measurement, key performance indicators

The main objective of the second workshop was to discuss and identify possible ways of measurement for the goal domains that had been defined in the first workshop. Each manager investigated possible operationalizations for one goal domain.

Managers presented their findings during the workshop. Subsequently, the most suitable indicators for each goal domain were chosen by consensus as key performance indicators.

4.3.C Workshop III: Ranges of indicators' levels

The third workshop mainly served to define the possible ranges of the indicators' levels. For each indicator, three values were agreed on by the management. They represented the lowest possible value, a neutral ("normal") value, and the best possible value. All values were chosen in a way that they were perceived as at the same time realistic and achievable by the whole group. While doing so, all indicators were scrutinized once more, leading again to some minor changes in the system of indicators.

4.3.D Workshop IV: Contingency functions

Contingency functions were developed using two different methodological approaches that are independent of each other. First, during the workshop, priorities of the indicators were defined for benefits (positive achievements) and harm (negative achievements) separately. Like this, two ranking lists were produced reflecting managers' consensus. Subsequently, each indicator's contribution to organizational performance was expressed as effectiveness score, showing its value for the overall organization. The indicator with the highest rank was set to 100 points and the indicators with the lower ranks were graded in relation to this value.

Second, managers were asked to attend a computer session, during which they individually answered 50 random paired comparisons of levels of indicators. Applying conjoint analysis [7, 8] using the commercially available software Alasca [9, 10], the indicators' usefulness for organizational performance was estimated. Conjoint analyses were calculated separately for airport (n = 4) and police managers (the last author being one of them) (n = 3), as well as averaged across all participants (n = 7).

In a last step, for each indicator, the contingency function that resulted from discussion was compared to the three contingency functions (airport managers, police managers, overall average) that resulted from the computer-based estimation. Differences were discussed, findings integrated, and last adjustments made to the contingency functions. Then, all participants agreed on the modified contingency functions by consensus.

4.4 Results

4.4.A Criteria and objectives

Fig. 1 shows the six goal domains and the 21 objectives that managers had agreed upon by consensus. Note that in this illustration, there is no hierarchy implied.

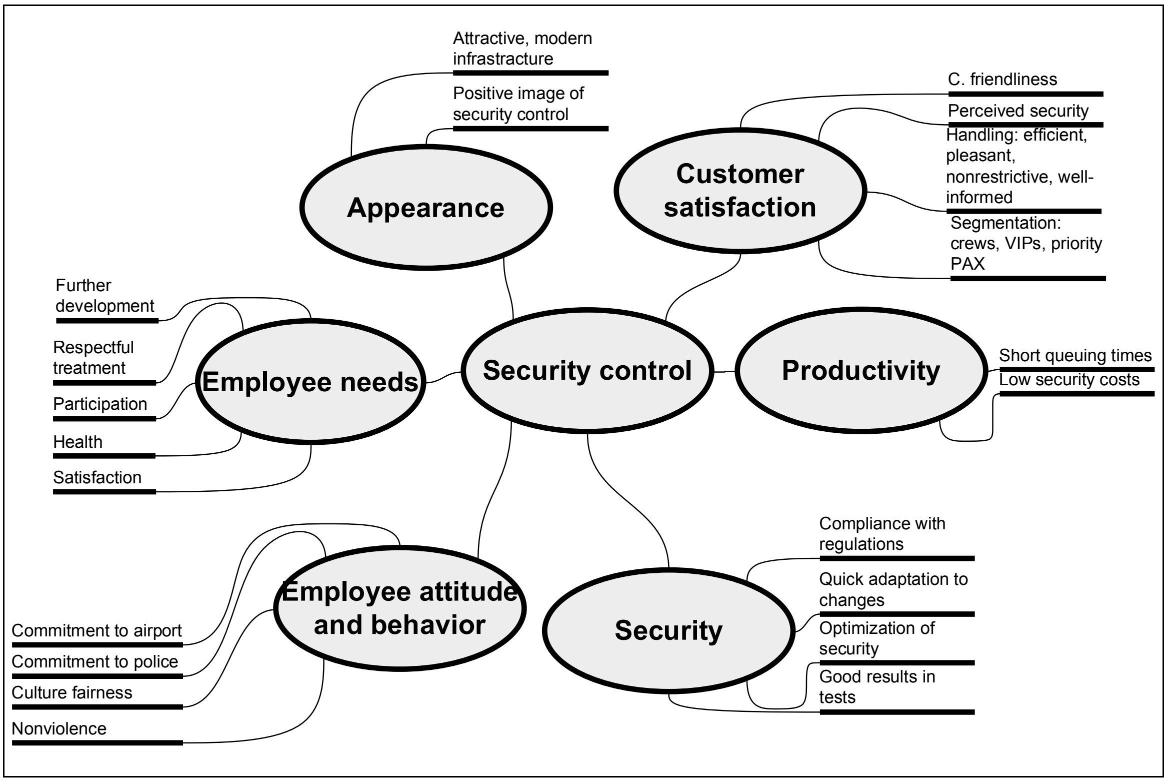

Fig. 1: Goal domains and objectives identified by the management.

4.4.B Key performance indicators

Table I

Key performance indicators.

No.	Goal domain	Indicator	Operationalization
1.1	Customer satisfaction	Satisfaction survey	Average of answers to 4 questions in independent survey
1.2	Customer satisfaction	Complaints	Overall sum of complaints documented
2.1	Productivity	Work min / # local PAX	Employee work minutes / number of local departing passengers
2.2	Productivity	Throughput	Mean throughput (PAX / h) of last 12 months during peak times
2.3	Productivity	Flight delays	Sum of flight delays caused by security control / total number of dep. flights
3.1	Security	Security quality	Average value of internal inspections and covert tests
3.2	Security	Security knowhow	Average value of SO certification and threat image projection data
4.1	Employee att. & behavior	Commitment survey	Average value of 2*5 questions on OCA
5.1	Employee needs	Satisfaction survey	Average value of 5 questions on employee satisfaction (JDS)
5.2	Employee needs	Time-lost-index	Average number of employee absences (h)
6.1	Appearance	Image survey	Average of answers on the security control’s image

Table I (see previous page) lists the six goal domains and their 11 respective indicators as well as their operationalizations. Note that, again, there is no hierarchy implied by the indicators' numbering. The hierarchy of the indicators is expressed solely by the contingency functions (see last paragraph of this section). The indicators shall be analyzed in reports that always cover a period of 3 months.

Goal domain 1, customer satisfaction, shall be measured with two indicators. The first one is the average value of 4 questions of a questionnaire. These are part of a yearly customer service survey carried out by an independent airport service benchmarking company. The second one is the sum of the written customer complaints about SOs' behavior. Seasonal differences (e.g., more families or passengers less familiar with the security control process are travelling during summer compared with other seasons) are known to be one external factor influencing security control [2]. In order to account for that, the sum of complaints is calculated in a rolling way, i.e., it is always calculated over the last 12 months and analyzed every third month for the ProMES framework. This procedure was adopted for all indicators that are measured in a rolling way (see below).

The discussion of goal domain 2, productivity, revealed that the common productivity indicators (e.g., throughput) reflect overall performance on a high and aggregated level. They are useful to managers, but the single employee has only a very limited influence on them and, therefore, might as well not be too interested in these data. It was decided that productivity shall be measured with three indicators: the first one reflects the ratio between revenue and expense. It denominates how many working minutes of employees are necessary to check one passenger. This value is calculated by the Airport Operation Manager by dividing the needed personnel hours by the number of local departing passengers of the respective time span. This value shall be calculated over the last 12 months (rolling). The second indicator, throughput, has been introduced by earlier research [2, 11]. In the present case, it is planned that data on throughput of peak hours (mean) and a clearly defined selection of control lanes only shall be analyzed. Analyses shall always contain data of the last 12 months (rolling). The third indicator, flight delays, consists of the sum of delayed flights (attributed to or caused by security control), divided by the total number of departing flights.

Goal domain 3, security, shall be measured with two indicators. The first one is an average of the ratings reached in internal inspections and covert tests carried out by the airport operator. The second one is an average of some selected results of the civil aviation authority's biannual screener certification for SOs and from threat image projection [12]. This value shall be calculated over the last 12 months (rolling).

Goal domain 4, employee attitude and behavior, is measured with one indicator that consists of 10 questions in total. These ten questions are made up of 5 items selected from the affective commitment scale (OCA) [13], translated to German by the authors. Each of those five items is asked twice, once regarding the police as direct employer of SOs and once regarding the airport as work environment.

Goal domain 5, employee needs, is measured with two indicators. The first one represents employee work satisfaction. It is measured using the general satisfaction subscale of the German version of the Job Diagnostic Survey (JDS) [14, 15], which consists of five items. Indicators 4.1 and 5.1 will be measured all three months by sending a questionnaire to a quarter of the workforce. Like this, every employee is invited once per year to complete a questionnaire. This questionnaire will feature the 10 questions of indicator 4.1 and the 5 questions of indicator 5.1. In addition to that, employees can provide reasons for their answers in an open format below the questionnaire items. Although this information is not rated in the ProMES system, it will be a crucial help in order to discover reasons if employee commitment and satisfaction increase or decrease. The second indicator of goal domain 5 is the so-called time-lost-index, which is the average number of hours that an employee is absent from work.

This is calculated by dividing the total number of work absence hours by the number of employees.

Goal domain 6, appearance, describes the appearance and the image of the security control, as seen by travelers and / or airport staff. It shall be measured with one indicator, which will consist of the average rating out of a certain number of questions that are still to be developed to this end. Interviews with travelers and / or airport staff shall be conducted by the airport operator.

The contingency functions for each of these 11 indicators are regarded sensitive information by the airport and are not further specified in this paper. For benefits (positive achievements), there are two most important indicators that share first priority. For harm (negative achievements), there is one indicator that is regarded to bear the most negative potential.

4.5 Discussion

The present study shows that it is possible to set up a ProMES system for security control on management level with reasonable investments of time and money. By agreeing upon contingency functions, existing goal conflicts can be addressed and discussed rationally on the basis of facts and numbers instead of personal opinions or subjective role interpretations. To be successful, it is necessary to involve all relevant stakeholders in order to develop the system. Because stakeholders have different interests, it can be helpful to have an independent facilitator supervise and moderate the process. Of course, this strongly depends on the work culture and the philosophy at the respective airport. This person should be well-experienced with ProMES, as the field of airport security and operation is rather complex due to the multitude of stakeholders, different interests, and interdependencies.

The seven participating managers each spent a total of 16.5 h for setting up the ProMES system on management level. In the whole developmental process, participants perceived discussion and exchange of different opinions as fair and constructive. This was probably due to the fact-based and consensus-oriented way of discussion provided by the ProMES methodology. On the interaction and atmospheric side of the project, the approach chosen led to only few incidents where different (conflicting) interests were being negotiated and to far more occasions where interests could be aligned, common goals for the managerial board be expressed, and methods for problem solving be applied. This resulted in broad acceptance of the solutions or decisions generated in each developmental step.

In both organizations, airport operator and police, there was one person who organized and prepared meetings and served as point of contact between entities and the independent facilitator. The first author did so for the police, the third author for the airport operator. The money invested consists of the independent facilitator's salary only as well as some minor expenses for professionalizing existing or developing new data collection tools.

The present application of ProMES in a security relevant domain is not the first one. ProMES has already been applied successfully at a Swedish traffic police unit [16]. However, the complexity of the situation at the airport had at first raised some concerns about the feasibility of such a project. The present study demonstrates that ProMES can successfully be installed on management level in security control. The resulting key performance indicators and the contingency functions are a useful tool to comprehensively monitor performance on a standardized and regular basis. Before that can happen, though, the exact properties of measurement have to be defined and reliability of measurement instruments checked. Data collection and analyses then have to be set up in practice, at times causing organizational effort.

From a theoretical perspective, it is obvious that a good performance indicator not only has to be reliable, but also valid and measured in a standardized way. This implies that, whenever possible, the viability of indicators should be critically questioned from time to time. After having set up a ProMES system with its respective indicators for the first time, reliability and validity analyses should be calculated as soon as enough data are available in order to optimize the system's practicability.

On setting up such a comprehensive measurement system for the first time, for some indicators, no data might be available (as it was the case for some indicators described above) so that the contingency functions have to be estimated based on theoretically presumed neutral, maximal and minimal values. As soon as enough data are available, the estimated contingency functions should be reassessed with the question in mind whether the functions are realistic or have to be adjusted. After having optimized and reassessed the contingency functions based on existing data where necessary, the functions should then be reconsidered in a further step with regard to each indicator's contribution to productivity. Indicators with a very flat slope (i.e., only marginal positive and negative impact on productivity) should be scrutinized critically, because such indicators do not contribute significantly to the overall performance of an organization. It has to be decided if they shall still remain in the system as key performance indicators, or if they are merely just performance indicators that are of lesser importance and need not be part of the system. Therefore, besides rather measurement-related considerations, the indicators should also be evaluated with regard to content.

However, ProMES is far more than just an instrument for measurement. A clear advantage of ProMES and its contingency functions is the fact that the importance or priority of each indicator is visually evident and therefore existing conflicts become apparent. By periodic measurement, trends and developments can be tracked and underlying reasons identified. Once those reasons are worked out, the company can install measures to either support those changes if they are positive, or to take countermeasures if they are negative. What is important, however, is that the organization as such can learn from the measurements and the visible trends. Because discussions between different stakeholders thereby are evidence and fact-based, constructive argumentation and the derivation of action plans that all participants feel committed to are fostered.

In order to achieve best possible effects in that direction, it is important that not only managers, but also employees know the basic principles of ProMES and have regular feedback sessions during which they can discuss results and decide on changes to be made. Managers agreed that there are some indicators that do not make sense to SOs because they are measured on a level of high abstraction and aggregation that has very little to do with their daily work. In this case, it is useful to let employees set up their own subsequent ProMES system within the six criteria defined by the management. They can thus find their own indicators which they can influence and sufficiently control in their daily work. This process might be more time and cost intensive than the first step that involved management level only. Nevertheless, we think that it is of crucial importance, at least in our case, to reach supervisors and SOs. The managers agreed that, by doing so, they want to foster participation and empowerment among employees. Furthermore, it was acknowledged that the employees themselves have a great influence on the quality of the final service that is offered to passengers.

Given the abundance of earlier studies on applications of ProMES, as well as the results of our study in a complex, security sensitive domain, we conclude that ProMES can be recommended as a methodology to set up a system for performance measurement and enhancement. We encourage international collaboration as well as exchange of ideas and experiences in this domain.

4.6 References

[1] Pritchard, R. D. (1990). *Measuring and improving organizational productivity: A practical guide.* New York, NY: Praeger.

[2] Wetter, O. E., Lipphardt, M., & Hofer, F. (2010). External and internal influences on the security control process at airports. In D. A. Pritchard & L. D. Sanson (Eds.), *Proceedings of the 44th International Carnahan Conference on Security Technology* (pp. 301-309). San José, CA: Institute of Electrical and Electronics Engineers.

[3] Ponemon Institute (2010). *Concerns about new airport screening procedures: U.S. survey of airline travelers.* Traverse City, MI: Ponemon Institute.

[4] Buzan, T., & Buzan, B. (1993). *The mind map book.* Harlow, England: BBC Active.

[5] Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard: Measures that drive performance. *Harvard Business Review, 70,* 71-79.

[6] Kaplan, R. S., & Norton, D. P. (1996). *The balanced scorecard: Translating strategy into action.* Boston, MA: Harvard Business School Press.

[7] Green, P. E., Krieger, A. M., & Wind, Y. (2001). Thirty years of conjoint analysis: Reflections and prospects. *Interfaces, 31,* 56-S73.

[8] Louviere, J. J. (1988). Conjoint analysis modelling of stated preferences. *Journal of Transport Economics and Policy, 22,* 93-119.

[9] Holling, H., Jütting, A., & Grossmann, H. (2000). *Alasca. Computergestützte Entscheidungs- und Nutzenanalyse. [Alasca. Computer-based analysis of decisions and benefit.]* Göttingen, Germany: Hogrefe.

[10] Höft, S. (2005). Tests und Tools: Alasca: Computergestützte Entscheidungs- und Nutzenanalyse. [Tests and tools: Alasca: Computer-based analysis of decisions and benefit.] *Zeitschrift für Personalpsychologie, 4,* 86-91.

[11] Wetter, O. E., Hofer, F., & Jonas, K. (submitted). Efficient and effective goal setting for short interventions.

[12] Schwaninger, A., Hofer, F., & Wetter, O. E. (2007). Adaptive computer-based training increases on the job performance of x-ray screeners. In L. D. Sanson (Ed.), *Proceedings of the 41st Carnahan Conference on Security Technology* (pp. 117-124). Ottawa, Canada: Institute of Electrical and Electronics Engineers.

[13] Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance and normative commitment to the organization. *Journal of Occupational Psychology, 63,* 1-18.

[14] Schmidt, K.-H., Kleinbeck, U., Ottmann, W., & Seidel, B. (1985). Ein Verfahren zur Diagnose von Arbeitsinhalten: Der Job Diagnostic Survey (JDS). *Psychologie und Praxis, Zeitschrift für Arbeits- und Organisationspsychologie, 29,* 162-172.

[15] Schmidt, K.-H., & Kleinbeck, U. (1999). *Job Diagnostic Survey (JDS – deutsche Fassung). In H. Dunkel (Ed.), Handbuch psychologischer Arbeitsverfahren [Handbook of psychological working methods] (pp. 205-230).* Zürich, Switzerland: vdf Hochschulverlag an der ETH.

[16] Pritchard, R. D., Culbertson, S. S., Malm, K., & Agrell, A. (2009). Improving performance in a Swedish police traffic unit: Results of an intervention. *Journal of Criminal Justice, 37,* 85-97.

5 General discussion

5.1 Further implications of the three studies

As many tasks in security are routine tasks coupled with a very low base rate of incidents, change is usually very much welcomed by employees. A frequently applied countermeasure against boredom in routine tasks is job rotation (e.g., Jorgensen, Davis, Kotowski, Aedla, & Dunning, 2005; Ortega, 2001; Schultz & Schultz, 2010). For tasks that do not require highly specialized knowledge, this is possible and could be advised. Some tasks, however, consist of highly specialized activities that make an interchange of functions much less practicable. For instance, a highly specialized task such as the screening of hold baggage can only be fulfilled by a group of specifically trained and certified SOs. They cannot be replaced on the spot by any other security personnel.

In these cases, it might be particularly fruitful to work with daily goals. Daily goals have the potential to provide some different views or foci on an identical task. The first study has demonstrated that goal setting can indeed be used in a similar way by supervisors in the everyday work setting at an airport's security control. These supervisors have had no prior experience with the use of daily goals. Moreover, they have not received any training in goal setting and communication. Nevertheless, positive effects could be found. However, in the long-run the use of this technique might present a challenge for supervisors. Having good, meaningful, and sensible daily goals ready over a longer period of time without too much repetition could be challenging indeed. Here, the supervisors' imagination, situational knowledge about what could be important on a certain day, as well as their in-depth knowledge of the possibilities of and variations on the task are required.

The first study analyzed the application of single goals in order to improve work performance in one particular aspect. In practice, though, requirements for employees are often more complex. For example, the open question of whether providing security and high throughput contradict each other has been of high practical relevance for many years. In science, this problem can be described as a classical speed-accuracy trade-off. The second study has given a first answer to that question. However, the answer is not a clear-cut "yes" or "no" because the occurrence of a trade-off depends on several conditions. The two most important ones are the individual employee's resources and capabilities, as well as the task's demands. It seems to be the balance between those two that determines whether a speed-accuracy trade-off occurs or not. The second study further analyzed whether two concurrent demands can be tackled by a dual goal, and if, by doing so, speed-accuracy trade-offs can be minimized. The result is promising and suggests that dual goals could also find their way into everyday practice in some particular situations. Moreover, the need for further research in the domain of multiple goals and their implications on work performance has been demonstrated.

The third study provided an overall framework for goal setting at a security control. It identified the main organizational objectives, defined key performance indicators accordingly, and developed operationalizations for measurement. With the present ProMES system, it can be assessed in which domain investments should be most useful to the organization. Moreover, the contents of this ProMES system can be communicated to the next lower level in the organizational hierarchy. Like this, transparency on expectations from management level to supervisor level can be achieved. As soon as supervisors know which goals are considered important by the management, and how the management prioritizes the different goal domains, they can tune their daily work accordingly. Using goal setting during their daily work, they have an influence on the optimization of the resources in their own sphere of action. If desired, supervisors could also set up their own ProMES system that is linked to that of the management, but that features other indicators which are modeled according to their own duties. Like this, a “nested” ProMES system could be developed that fits logically into the ProMES system of the management representatives. Thus, a holistic system of goal setting using the ProMES methodology could be set up in an organization, which would be highly recommended as a next step.

Setting up a ProMES system with employees that are unfamiliar with the approach is also a challenge regarding communication. Talking about performance and performance measurement can very quickly be perceived by employees as something dangerous. Indeed, one can imagine that ruthless management representatives could set up such a system in order to exploit employees. Particularly in combination with very low base pay and financial penalties for employees that are below expectations, such a measurement system can pose a threat to employees. Employees will react to this in a certain way, which might include having a narrow focus that neglects non-goal areas, a rise in unethical behavior, distorted risk preferences, and reduced intrinsic motivation, which could lead to corrosion of the organizational culture (Ordóñez, Schweitzer, Galinsky, & Bazerman, 2009a). The pointing out of these potentially dangerous effects of goal setting by the above authors has started a controversy. Locke and Latham (2009) replied to the article mentioned above, which again triggered another reply by Ordóñez, Schweitzer, Galinsky, and Bazerman, (2009b).

However, Klaus-Helmut Schmidt (personal communication, May 28, 2010) stressed that participation, empowerment, and motivation are the key concepts behind ProMES, which could be described as some sort of underlying “philosophy” to the method. He mentioned that ProMES projects where this had been neglected usually turned out unsuccessful. One key idea of ProMES is that individuals and teams optimize themselves by analyzing performance feedback and the underlying causes for performance that is better or worse than expected. Subsequently, they take action and can evaluate the success of the measures taken in the next performance feedback report. Optimizing team performance is about employees helping and supporting each other and thus optimizing “from within”. Getting better as a team and subsequent appraisal and recognition by supervisors is motivating.

ProMES can also serve as a platform for developing team spirit. It can be expected that teams with high cohesion, mutual trust and support show Organizational Citizenship Behavior (Organ, 1988, 1997) towards the other team members and towards the organization. Thus, not only the employees, but also the organization as a whole benefits from a ProMES system that is set up in a sensible way and that is accepted by the employees. This requires managers with a high degree of responsibility, mutual trust between the different levels of the hierarchy inside the organization, and a system of performance measurement, management, and appraisal that is clever, sensible, and socially acceptable. To convince employees that this is the plan and to motivate them to participate in setting up the ProMES system is a great deal of work and requires good communication skills and strategies.

5.2 Outlook into the future of airport security research

The general introduction has aimed at familiarizing the reader with the rise and development of psychological issues in airport security research. The main body of this thesis, consisting of the three studies, has introduced methods and principles from work and organizational psychology into the field of airport security. As has been demonstrated, airport security was subjected again and again to severe changes, which have usually been triggered by terrorist attacks. As the airline business in general and airport security in particular are subject to quick changes, this field is particularly interesting but also challenging for researchers. There is no point in doing research on certain conditions that might have changed unrecognizably before the data of the study are analyzed and the manuscript is written. Given the quick-paced nature of this business, many more developments can be expected to occur during the next years.

5.2.A *Innovation domain I: technology*

The first domain where such developments could take place is technology. Improvements in technology can be of different nature and on different levels: At the research level, scientists can develop completely new technologies (e.g., Pulsed Fast Neutron Analysis or Thermal Neutron Analysis for detecting items that are not allowed aboard an aircraft, see Seidenstat, 2009) that can subsequently be implemented by the security industry. At the regulator level (in Switzerland, this is the FOCA[1]), performance standards for devices to fulfill for approval can be made more stringent. At the manufacturer level, currently available and implemented technology can be further improved and upgraded. Already existing technologies that are not yet deployed (e.g., security scanners[2] in Switzerland), could further be improved together with researchers so that they reach readiness for deployment. At the operator level, airports, police, and private security companies can invest into better devices that offer performance above the minimal standards.

For all new equipment, rigorous testing is necessary. It has to be made sure that devices fulfill the performance standard prescribed by the regulator (type approval). Moreover, continuous functional checks during everyday use have to prove that the device is still operating with unimpaired performance. For type approval, but especially for later procurement of devices by airports, operational tests are of utmost importance. The objective of operational testing is to reveal the impact of a new device on operational issues such as required operator training, acceptance by passengers, space and maintenance requirements, integration into existing processes, and processing speed.

---

[1] The FOCA (Federal Office of Civil Aviation) is "responsible for monitoring civil aviation in Switzerland and aviation development. It is responsible for ensuring that civil aviation in Switzerland has a high safety standard and one that it is in keeping with sustainable development" (FOCA, 2011).

[2] "Security Scanners" (also known as "body scanners") are devices that are used to check if items are carried on the body or in garments that are not allowed in the cabin of the aircraft.

5.2.B Innovation domain II: human factors

The second domain where developments could take place is the human factors issues. One way of improvement lies in the optimization of existing concepts and procedures. In pre-employment assessment, reliable and valid assessment centers, taking into account all of the duties of SOs (not just x-ray screening) could be developed. In practice, approaches exist already, but they have to be evaluated, improved, and published. In training, syllabi have to be continuously updated and improved. For example, correct operation of a new device as well as subsequent interpretation of results and alarm resolution have to be learned and practiced by SOs so that they are fit for work and ready for taking correct actions in case of an incident. Improvement and updating can refer to the contents of the training as well as to methods of and media used for teaching.

Regarding evaluation and performance measurement of SOs, many approaches have been described in this thesis (e.g., TIP, throughput, passenger density, covert tests). Today, each approach has certain shortcomings that could be improved. For fields where no approach exists up to present, new ways of performance measurement could be developed and introduced. In the broad field of leadership, the work has just begun with the studies of this thesis. In practice, the findings about goal setting have to be taught to new supervisors so they can make use of the available knowledge during their everyday work. In parallel, further knowledge about leadership issues in security has to be gained.

A domain that is receiving growing attention is the study of suspicious behavior. It is believed that cognitive and emotional clues, facial expressions, and behavioral-impression management clues (Frank, Maccario, & Govindaraju, 2009) could be used to identify suspicious persons before they commit a crime. Bättig, Frey, and Hofer (2011) carried out a study that showed that recognition of thieves by studying their behavior prior to the theft is possible. They revealed that policemen were significantly better at early identification of thieves than laymen. It can be expected that the findings of this study about theft can be used as a starting point for later generalization to early identification of terrorists. If terrorists or individuals with malicious intents could be spotted reliably, this would theoretically allow for an alleviation of the stringent security control measures for "normal" passengers.

5.2.C Innovation domain III: organizational systems and responsibilities

Many airline representatives call for a risk-based approach for shaping the future of security control. They claim that risk assessment (Dillingham, 2003) allows for a grouping of passengers or baggage into several categories that shall be checked in different depth; an approach that would save costs (e.g., Nie, 2011; Nie, Batta, Drury, & Lin, 2009). Some of the crucial properties of this approach, however, remain unclear: Up to the present, it is not known if risk assessment allows for a reliable selection of passengers or baggage that pose high or low threat risk. Moreover, this procedure is highly questionable from an ethical point of view. An ethical question that is certainly raised if this approach is applied to passengers is how much it complies with nondiscrimination and equal treatment of human beings. A related approach that is probably more fruitful is randomization. Randomizing certain security checks is a good measure for preventing adversaries observing and exploiting patterns and processes of checks (Pita et al., 2009).

Another hot topic that is likely to receive growing attention in the future is the insider threat. This term refers to people within an organization that disclose confidential information (espionage), intentionally misuse systems, or commit sabotage. Most research on this topic was done in information technology (e.g., Moore, Cappelli, & Trzeciak, 2008). As the world's dependency on information technology is steadily increasing, this topic becomes more and more important. In the security domain, information technology is just as present as in any other branch of the economy. Therefore, it can be expected that the insider threat will become an important topic of security research in the future.

A variety of human and organizational factors might decrease the likelihood of occurrence (threat risk) of as well as the vulnerability for an insider attack. As Felix Walz, head of the staff branch of the airport police of Kantonspolizei Zürich stated, well-trained supervisors with a keen perception, who interact frequently with their subordinates, should be able to feel when something changes or is wrong in one of their subordinates (personal communication, August 2010). A good organizational climate with mutual respect, support, and care is also likely to contribute to a low incidence of insider crime in an organization. Moreover, good conditions of employment and work (e.g., decent salaries, acceptable working times, healthy working conditions) might further reduce the likelihood of insider attacks. However, the above facts are mere speculations and will have to be the subject of future studies.

Some legal and political issues, for example the question of which entity is in charge of security control, are also regarded as an influence on the provided quality level of a security control: In the United States, prior to the 9/11 attacks, private security companies had been carrying out security control, whereas after 9/11, this responsibility was taken over by the government (Kirk, 2003) in order to establish a higher level of security (Seidenstat, 2009). In most European countries, security control at airports is still carried out by private security companies, whereas at Zurich Airport, the police (Kantonspolizei Zürich) are mandated by law (Kanton Zürich, 2008) to carry out security control. Whether privatization is to be favored or not depends on viewpoint: Airport-related authors in general rather favor privatization (e.g., Mew, 2005) because it offers more flexibility and lower costs, whereas other sources state that federal authorities are doing the job rather well (e.g., Gkritza, Niemeier, & Mannering, 2006; Seidenstat, 2009).

5.2.D Innovation domain IV: interdisciplinary approaches

Although discussed separately above, technology and human factors should also be studied with interdisciplinary approaches regarding their interplay. Traditional fields in that domain are ergonomics and human-machine-interaction. Operational testing of equipment, as described above, is basically one of the purest forms of analyzing human-machine-interaction. By doing so, the strengths, weaknesses, and the performance output of the human-machine-system as a whole are measured. The holistic view and analysis provided by operational testing in the field is without doubt closest to the real performance output that can be expected in everyday use. Recently, the focus has broadened even more as sociologists and ethicists have entered the area. Ethics cover not only human-machine-interaction, but even add the society with its norms to the picture. Up to the present, ethicists have been particularly active in studying the impact of security scanners on society (e.g., Nagenborg, 2011; Traut, Nagenborg, Rampp, & Ammicht Quinn, 2010).

5.2.E Funding and the "embedded" approach

In the early days, funding for security research in Europe was scarce since applied research and development cannot compete with basic research for federal or state funding. An often heard argument was that the economy should pay for advances in that domain. In some cases, this had indeed happened, although the economy basically only invests money if there are good chances for a return on investment. In other cases, it was possible to get mandated by the regulator (e.g., for developing certification tests for x-ray screeners). Recently, the recognition by some governments and the public, as well as the importance attributed to applied research and development has increased. As a result, there are now several state funding programs for security research in Europe (e.g., by the European Union, the German Federal Ministry of Education and Research, and, as of 2012, the FOCA).

Of course, funding is a necessary precondition for fruitful research. However, in the security domain, projects funded in this way have met with mixed success until now. Some might have yielded scientifically interesting results, but have not added much practical value nor led to many improvements in security. The problem is often that researchers at universities are physically separated from practical problems, and may therefore focus on issues peripheral to the problems encountered during daily operation of security control. Moreover, they usually have highly restricted access to security sensitive information or to security personnel, which would enable them to test new concepts. If research on security sensitive information can be done, it is usually done by state labs and cannot be published, which is of course not very attractive to researchers. With our new approach of research and development that is embedded in police structures, many of these problems can be avoided (although different ones might come up). It is our goal to contribute to shaping the future of security research by providing what others cannot provide and by fostering interdisciplinary approaches and collaboration between established and prestigious academic research and the real-life setting of airport security control.

5.3 A discussion of the necessity of a psychological security theory

In the General Introduction (chapter 1.1), a historical overview of the developments in airport security research was given. The three studies of this thesis can be seen as a continuation of this line of research (chapter 1.3). On taking this knowledge into account, it becomes evident that all the developments up to the present consisted of applications of existing psychological theories, concepts, and methods in the newly opened field of airport security. All of the underlying psychological theories (see chapter 1.4) are general psychological theories that do not focus on security in particular.

I would like to argue that the field of airport security has not yet been able to exert a large and significant impact on the body of psychological theories. Until now, the main value of the studies in airport security to psychology as an academic subject consisted of the practical insights provided: A deeper understanding of processes, requirements, and practices in security has been reached. The study of airport security has also been a helpful and interesting new field for (re-)testing and application of existing psychological theories, approaches, and management systems. By doing so, contributions to existing theory and broadening of the available psychological knowledge came almost as a byproduct and were not of large impact.

There are two possible reasons for this lack of contribution of studies in airport security to the body of psychological theories until now: The first reason might be that the subject is simply still too young to have made a significant theoretical contribution. If a new subject is opened, its properties and mechanisms have first to be studied and understood by its researchers. They have to become familiar with the nature of their new subject before they can condense their knowledge into new theories. If that should be the case with airport security research, new theories that significantly contribute to psychology as an academic subject can be expected in the future. The second reason might be that airport security research is basically not able to contribute significantly at all. It appears possible that no new theories are necessary for psychological aspects in airport security. In that case, the field might stay interesting nevertheless for applied research that helps practitioners improving current processes and applications. The academic interest in the subject, however, might decrease over time given that case.

These thoughts lead us inevitably to the question under which circumstances it would be necessary, useful, and fruitful to create a new theory for psychological aspects in airport security, and in which cases this could be deemed unnecessary. In my view, a new theory would be beneficial in three cases: (1) if it explained phenomena in airport security in a more precise and detailed way than the available psychological theories, (2) if it explained phenomena in airport security in a simpler way than the available psychological theories (e.g., if some things worked differently in airport security than in other fields, many exceptions and additional statements might have to be added to existing theories in order for them to remain valid), and (3) if it explained properties and relations between properties in order to predict outcomes that are not at all covered by existing psychological theories.

In his much celebrated lecture “truth, rationality, and the growth of scientific knowledge”, Popper (1963) formulated six types of cases in any of which a new theory might be seen as closer to the actual facts than an earlier theory. A comparison of my three statements of the previous paragraph to Popper’s ideas reveals some parallels. I perceive his first case, that a new theory would be fruitful if it made more precise assertions that should stand up to more precise tests, to be of identical notion as my first statement. His second case, that a new theory would be fruitful if it explained more facts, can be seen as corresponding with my third statement. His third case, that a new theory would be fruitful if it explained facts in more detail, can be seen as included in my first statement.

Popper’s fourth case, that a new theory would be fruitful if it passed tests that the available theory has failed to pass, is without doubt important to science in general. For practitioners working in the field, it might seem to be of lesser importance, though. Practitioners’ most important goal is usually to apply knowledge to the field, whereas improvements over existing psychological theories (a new theory passing tests that an old one has failed) are probably no major concerns, as long as the old theories remain valid and useful. Of course, every new theory has to be tested for validity and should, in general, have some advantages over the old one. Popper’s fifth case, that a new theory would be fruitful if it suggested new experimental tests, as well as his sixth case, that a new theory would be fruitful if it unified or connected hitherto unrelated problems, might be useful for both practitioners and scientists. The fifth case might suggest that a new theory could bring about new (i.e., better or more precise) ways of measurement, whereas the sixth case hints at the possibility that by combining problems or parts of existing theories, fruitful results can be obtained.

In political science, security theories exist already. They usually refer to the overall security processes going on in nations, internationally, and globally (e.g., Booth, 2007; Haftendorn, 1991; Tang, 2004). These theories are on the highest hierarchical levels as states already feature an extremely high degree of process aggregation. Security issues at the state level are the outcome of an integration of security issues of all the subordinate hierarchies (e.g., the individual behavior of the people living in this state, the behavior of the groups, the teams in organizations, the organizations as a whole, the political parties, systemic aspects in politics, law, society, etc.). Security theories in political science are without doubt interesting for a study of the behavior of states, several states, international organizations, and the global world regarding security issues. In other words, security is defined in these theories as a rather global concept. In contrast, a psychological security theory would probably be set on a lower hierarchical level. It could target human (i.e., individual, dyad, group, and organizational) behavior in the security domain. Thus, it would in particular allow for a more fine-grained analysis and description of the aspects in security that are caused, shaped, or influenced by humans.

In the same lecture, Popper (1963) further stated that the development of new theories is actually the driving force of science. New theories and the critical examination of their predictions drive the development of new experiments and observations, which can in turn lead to interesting discoveries. He further hinted at the fact that science starts from problems, and not from observations. Thus, a careful scrutiny of the existing problems in airport security, the analysis of the properties involved in the problems, and an idea about how they relate to each other might eventually lead researchers in airport security to construct a theory that solves one or several existing problems. By deriving new techniques or interventions from the new theory, some problems might also be solved in practice.

Up to the present, some problems are identified and known to practitioners and scientists. Scientific knowledge has contributed to several single solutions that have proved effective and efficient. However, an overall framework has not yet been developed. By making use of the embedded applied research and development paradigm, researchers might be able to set up theories that, by providing an overall framework for human factor aspects in airport security, contribute to psychology as a scientific subject and the improvement in practical security alike. In my opinion, this will be one of the major future challenges in airport security research.

However, progress takes place in small steps; with every solution to a problem, new problems or at least questions arise. In my view, this is particularly evident in airport security research. Scientists have addressed many problems, and solutions have been implemented in practice. However, terrorists could be described as some sort of “intuitive researchers”, who, like their adversaries on the “good” side, will always establish new theories and activities, too. Terrorists strive, at least unconsciously, for broadening their understanding of security systems. It can be assumed that they develop subjective and maybe “naive” theories about weaknesses of the security processes and how those can be exploited. Like this, the “good” and the “bad” forces could be seen as opposing poles that drive the spiral of research and development in airport security. Unfortunately, in many cases in the history of airport security, the “bad” forces were on the leading edge and the “good” forces rather answering the challenges posed by their adversaries. It is time for making a step ahead.

5.4 References

Bättig, F., Frey, C., & Hofer, F. (2011). ASPECT: Analysing suspicious persons and cognitive training. *Kriminalistik, 10,* 641-647.

Booth, K. (2007). *Theory of world security.* Cambridge: Cambridge University Press.

Dillingham, G. L. (2003). Aviation security: Progress since September 11, 2011, and the challenges ahead. In J. Zellan (Ed.), *Aviation security: Current issues and developments* (pp. 1-22). Hauppauge: Nova Science Publishers.

FOCA (2011). *Duties of the Federal Office of Civil Aviation.* Retrieved November 30, 2011, from http://www.bazl.admin.ch/org/aufgaben/index.html?lang=en.

Frank, M. G., Maccario, C. J., & Govindaraju, V. (2009). Behavior and security. In P. Seidenstat & F. X. Splane (Eds.), *Protecting airline passengers in the age of terrorism* (pp. 86-106). Santa Barbara: ABC-CLIO.

Gkritza, K., Niemeier, D., & Mannering, F. (2006). Airport security screening and changing passenger satisfaction: An exploratory assessment. *Journal of Air Transport Management, 12,* 213-219.

Haftendorn, H. (1991). The security puzzle: Theory-building and discipline-building in international security. *International Studies Quarterly, 35,* 3-17.

Jorgensen, M., Davis, K., Kotowski, S., Aedla, P., & Dunning, K. (2005). Characteristics of job rotation in the Midwest US manufacturing sector. *Ergonomics, 48,* 1721-1733.

Kanton Zürich (2008). *Gesetz über den Flughafen Zürich (Flughafengesetz).* [Law on Zurich Airport (Airport law).] Zürich: Author.

Kirk, R. S. (2004). Selected aviation security legislation in the aftermath of the September 11 attack. In J. Zellan (Ed.), *Aviation security: Current issues and developments* (pp. 45-64). Hauppauge: Nova Science Publishers.

Locke, E. A., & Latham, G. P. (2009). Has goal setting gone wild, or have its attackers abandoned good scholarship? *Academy of Management Perspectives, 23,* 17-23.

Mew, K. (2005). Airport security screening: Privatize or federalize? *Public works management & policy, 10,* 3-9.

Moore, A. P., Cappelli, D. M., & Trzeciak, R. F. (2008). *The “big picture” of insider IT sabotage across U.S. critical infrastructures.* Pittsburgh: Carnegie Mellon University.

Nagenborg, M. (2011). Körperscanner. [Body scanners.] In M. Maring (Ed.), *Fallstudien zur Ethik in Wissenschaft, Wirtschaft, Technik und Gesellschaft [Case studies about ethics in science, economy, technology, and society]* (pp. 236-242). Karlsruhe: KIT Scientific Publishing.

Nie, X. (2011). Risk-based grouping for checked baggage screening systems. *Reliability Engineering and System Safety, 96,* 1499-1506.

Nie, X., Batta, R., Drury, C. G., & Lin, L. (2009). Passenger grouping with risk levels in an airport security system. *European Journal of Operational Research, 194,* 574-584.

Ordóñez, L. D., Schweitzer, M. E., Galinsky, A. D., & Bazerman, M. H. (2009a). Goals gone wild: The systematic side effects of overprescribing goal setting. *Academy of Management Perspectives, 23,* 6-22.

Ordóñez, L. D., Schweitzer, M. E., Galinsky, A. D., & Bazerman, M. H. (2009b). On good scholarship, goal setting, and scholars gone wild. *Academy of Management Perspectives, 23,* 82-87.

Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington: Lexington Books.

Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance, 10*, 85-97.

Ortega, J. (2001). Job rotation as learning mechanism. *Management Science, 47,* 1361-1370.

Pita, J., Jain, M., Ordóñez, F., Portway, Ch., Tambe, M., Western, C., … Kraus, S. (2009). Using game theory for Los Angeles airport security. *AI Magazine, 30,* 43-57.

Popper, K. (1963). *Conjectures and refutations.* London: Routledge.

Schultz, D. P., & Schultz, S. E. (2010). *Psychology and work today.* Boston: Pearson.

Seidenstat, P. (2009). Protecting airline passengers: An overview. In P. Seidenstat & F. X. Splane (Eds.), *Protecting airline passengers in the age of terrorism* (pp. 3-12). Santa Barbara: ABC-CLIO.

Tang, S. (2004). A systemic theory of the security environment. *The Journal of Strategic Studies, 27,* 1-34.

Traut, A., Nagenborg, M., Rampp, B., & Ammicht Quinn, R. (2010). Körperscanner: Sicherheiten und Unsicherheiten [Body scanners: Securities and insecurities]. *Forum Kriminalprävention, 1,* 14-20.

6 Postscriptum

6.1 Statement of main contributions

The first study was designed, carried out, analyzed, interpreted, and written by myself. BSc Mirjam Baur assisted me in data collection; Prof. Dr. Klaus Jonas revised my manuscript. The first experiment of the second study was designed and carried out by Profs. Jürgen Wegge and Klaus-Helmut Schmidt. The data were analyzed and interpreted by me, supervised by Prof. Jürgen Wegge. The second experiment was designed, carried out, analyzed, and interpreted by myself. For data collection, I was again assisted by BSc Mirjam Baur. I wrote the entire manuscript, which was subsequently revised mainly by Prof. Jürgen Wegge. The third study was designed, carried out, and written by myself. Dr. Hartwig Fuhrmann conducted the workshops with the management representatives. In organizational matters, I was assisted by Dr. Franziska Hofer. Both Dr. Franziska Hofer and Dr. Hartwig Fuhrmann revised my manuscript.

6.2 Acknowledgments

I would like to thank Prof. Dr. Klaus Jonas for supervising me in the best way one can imagine. Further thanks go to Prof. Dr. Alexandra Freund for her kind support as second referee for this thesis. I would further like to thank my employer, Kantonspolizei Zürich, for offering me the possibility to work in and further study the interesting and challenging field of airport security. I am grateful to Oblt Felix Walz and Dr. Franziska Hofer who were the main responsible persons for supporting my PhD at work. Personal thanks go to Dr. Emma Slack for all the miracles we discover and share together.

6.3 Curriculum vitae

Olive Emil Wetter was born in Bern, Switzerland, in 1980. He first studied music at the Bern University of the Arts, where he graduated with a teacher's diploma (equivalent to a BA) and, subsequently, with a performer's diploma from Prof. Bruno Canino's master class in 2003. He has been working as a piano teacher and performing artist in different European countries since. His work is also reflected in newspaper articles, radio broadcasts, and CD recordings.

In 2003, he started his studies of Psychology, Psychopathology, and English Literature at the University of Zurich. While studying, his interest in airport security issues was awakened. He received his MSc in 2008 for his thesis entitled "Using TIP data for assessing the impact of image-based and human-based factors on detection performance in x-ray screening". During his PhD studies with Prof. Klaus Jonas, he tried to provide a new perspective to airport security by introducing methods and principles from work and organizational psychology.

Since 2007, he works for Kantonspolizei Zürich, Airport Police, and is a founding member of its research and development team. He published various articles and held several talks on aviation security issues. In 2011, he chaired a symposium on security research in Switzerland at the 12th congress of the Swiss Society of Psychology. His teaching experience includes lectures at the Universities of Zurich and Bern, as well as courses for Kantonspolizei Zürich. He works as an expert for the European Civil Aviation Conference (ECAC), the Association of German Engineers (VDI), as well as the Association of Swiss Security Service Companies (VSSU).